Spring 1-1-2011

Caring Satisfactionism: A New Theory of Personal Welfare

Jason Hyde
University of Colorado at Boulder, jhyde0@gmail.com

CARING SATISFACTIONISM:

A New Theory of Personal Welfare

by

JASON HYDE

B.S.B.A., University of Missouri, 1992

M.B.A., Baylor University, 1993

J.D., University of Texas, 2000

B.A., University of Texas, 2005

M.A., University of Colorado, 2008

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Philosophy

2011

This thesis entitled:
CARING SATISFACTIONISM: A New Theory of Personal Welfare
written by Jason Hyde
has been approved for the Department of Philosophy

Christopher Heathwood

Robert Hanna

Alastair Norcross

Michael Tooley

Benjamin Hale

Date_____

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

ABSTRACT

Hyde, Jason (Ph.D., Philosophy)

CARING SATISFACTIONISM: A New Theory of Personal Welfare

Thesis directed by Assistant Professor Christopher Heathwood

Some lives go better than others. On this fact there is virtually no disagreement. If that is true, then what makes it so? Answers to this question are theories of personal welfare. This dissertation provides such a theory that claims, roughly, that a life goes better for the person who lives it to the extent he gets what he cares about getting and he believes he is getting those things.

This dissertation is structured as follows. The Prologue properly frames and details the importance of the issue of personal welfare. Chapter One examines two of the main types of personal welfare theories. Objections to those theories are raised, and some general principles of welfare are formulated in response to those concerns, including that desires play at least some role in the correct theory of welfare. Chapter Two takes a preliminary look at desire theory—the third and final main type of personal welfare theory. It considers and rejects the possibility that something in addition to desires affects welfare before taking stock of the challenges that face any desire theory. Chapter Three begins by considering whether a unified theory of welfare for all beings exists before rejecting this idea. The main focus of the chapter, however, is putting in place the foundations of the proposed theory as found in the works of Harry Frankfurt. Frankfurt's works on a variety of topics are covered in great detail because part of the problem in the debate over desire theory stems from the failure to view desires in their proper and broader context as an integral part of the human psyche. Chapter Four considers some popular

forms of and objections to desire theory in order to collect the remaining conceptual tools required to build the correct theory of personal welfare. Chapter Five explains this theory in full, extols its virtues, demonstrates how it resolves the two most difficult objections facing any desire theory, and shows how it resolves a deep conceptual confusion within desire theory. Finally, the Epilogue details the meaning of life and shows how Caring Satisfactionism is perfectly consistent with it.

DEDICATION

To Mom and Dad, Spike and Bullet, and Michel

ACKNOWLEDGMENTS

I began working on the ideas that eventually turned into this dissertation in early 2006. My friend, Barrett Emerick (thank you, Barrett), let me join in on an independent study class he had scheduled with (what turned out to be) my dissertation supervisor, Chris Heathwood. We read Derek Parfit's thick, dense, and brilliant *Reasons and Persons* cover-to-cover that spring. A few pages from the end of the book, I came across the first serious, sustained discussion of one of the reasons I wanted to do philosophy in the first place: What makes someone's life go best?

My plan was to eventually turn the paper I wrote for the independent study into my dissertation. And that is sort of how it turned out. My first stab at this topic of personal welfare was to create a new version of a desire-satisfaction theory in which the salient feature was the claim that our desires about our desires are what ultimately determines how well a life goes for the person who lives it.

That paper was okay for a first attempt, but I had a lot to learn about this topic. After turning in the paper for the independent study, I turned in a different version of it to the department for the first of two required diagnostic papers. The single most critical substantive spark for this project appeared in the comments that I received back. One of the anonymous reviewers (thank you, anonymous reviewer), in what seemed to be an afterthought, wrote something to the effect of, "Hey, you should check out Harry Frankfurt's work on this desires about desires stuff."

I was completely unaware of Frankfurt's work in this area at the time, but the light bulb went on for me once I dug into his oeuvre. Although I disagreed with much of what Frankfurt argued, I could see that the seeds for the theory I was trying to perfect were there. Over the course of the last four years, however, I came to see that many of my ideas were untenable in the face of Frankfurt's arguments. Quite simply, he has done a masterful job of giving a wonderfully nuanced account of most of the issues relating to our experience of ourselves and how that relates to the things we care the most about. I have already had the opportunity to thank him via email, but he certainly deserves another one here. So thank you—Harry Frankfurt—both I and this project are much richer for having had the benefit of your insights.

The other great intellectual debt I owe is to my supervisor, Chris Heathwood. Chris studied under Fred Feldman, another great philosopher who shows up a lot in the coming pages, and wrote his dissertation on personal welfare, too. He learned his craft exceedingly well and has been extraordinarily generous with both his time and his philosophical talents. If I were to acknowledge his influence on every page of this project that merited it, then it may start to look like I was merely the figurehead for a ghost-writing project Chris undertook for unknown reasons. So thank you—Chris Heathwood—you made me see, time and again, the force of the arguments, both large and small, that (hopefully) kept me on the right path.

As even a cursory examination of this project would reveal, I owe a great deal to many other philosophers as well. In particular, David Hume and

Friedrich Nietzsche taught me, apart from numerous substantive lessons, that good philosophy need not be inaccessible or boring. I have tried to heed these lessons by making this dissertation accessible and interesting; I hope I have succeeded to some degree. I also owe a great deal to a number of philosophers who, particularly in the past 30 years or so, have been working diligently on the topic of personal welfare. Philosophy is hard, but not nearly as hard as it would be without standing on the shoulders of great philosophers. It is a humbling fact that most of the good parts of this project, if there are any, are due to others while the bad parts of this project are mine alone.

I would also like to thank the other members of my dissertation committee—Michael Tooley, Bob Hanna, Alastair Norcross, and Ben Hale—and the CU Philosophy department in general. You gave me a chance to fulfill a dream and made it a very rewarding and enjoyable process to boot. Other than the temperature, I would have had trouble designing a better set of circumstances in which to get my Ph.D.

I would also be terribly remiss if I did not thank my mom, Donna Budnick, and my dad, Jerry Hyde, for all they have done to get me to this point. In many ways—some very obvious and some much less obvious—I would not be anything I am without your love and support. Similarly, Spike and Bullet have been as good of friends as I could have ever hoped for. I would not be where I am without you two, either. Also, a former student turned friend of mine, Mike Ebben, helped to keep me sane(?) throughout the writing of this dissertation with a weekly backgammon game during which—between

repeated amusing endorsements of the wonders of the experience machine (a contraption you will meet shortly)—he let me bounce off him other ideas relating to this project.

Finally, I often feel as if I will never be able to repay some of the debts I owe. I have never been more sure of that than as it relates to Michel. I am not even sure how to start to thank her, much less to attempt to repay her. So here goes nothing. She moved to Colorado to be with me when I embarked on this adventure. She hates the cold. She billed (long) hours at a law firm to pay the vast majority of our bills. She hates paying attention to time. She took the bus to Denver for a job for a while. She hates commuting. She complained about some of those things, but far less than she was entitled to and far less than I would have. And she didn't complain about my earning less money, working less hard, having more fun, or any number of other things that would have been hard for me not to comment on had the roles been reversed. When it came time to write my dissertation, she was genuinely encouraging all the time. But that was the least of her contributions to this project. For not only did she type many long passages from many books into my notes, she typed this entire project. Yes, I wrote this by hand using up 8+ Blue Papermate Profile pens and about as many yellow legal pads. The kicker here is that my handwriting is terrible. Really, really terrible. Yet she deciphered my hieroglyphics with truly stunning accuracy, typed it all, and did not complain. She also did more grunt work on this than she probably wants to be reminded of. Oh, and did I mention that she is something of a savant when it comes to

all aspects of editing? Yes, I think she might know all of the rules, and she is also extraordinarily patient and meticulous (I guess an Ivy League education is good for something after all). I could go on, but suffice it at this point to say that if there were a way she could support me through this process, then that is what she did. Honestly, there are no more ways she could have supported me. And not one genuine word of complaint about this grand adventure of mine. In short, there is no chance that I would have finished this dissertation without Michel. Do you think I will be able to repay her? I am laughing thinking about the answers I might get to that question. So thank you, Michel, thank you. You are The Giving Tree. I just hope I am not the boy.

TABLE OF CONTENTS

What gives this mess some grace unless it's fictions
Unless it's licks, man
Unless it's lies or it's love?

—Okkervil River

PROLOGUE: WHO CARES?

> One sticks a finger into the ground to smell what country one is in; I stick my finger into existence—it has no smell. Where am I? What does it mean to say: the world? What is the meaning of that word? Who tricked me into this whole thing and leaves me standing here? Who am I? How did I get into the world? Why was I not asked about it, and why was I not informed of the rules and regulations but just thrust into the ranks as if I had been bought from a peddling shanghaier of human beings? How did I get involved in this big enterprise called reality? Why should I be involved? Isn't it a matter of choice? And if I am compelled to be involved, where is the manager—I have something to say about this. Is there no manager? To whom shall I make my complaint? After all, life is a debate—may I ask that my observations be considered? If one has to take life as it is, would it not be best to find out how things go? . . . Will no one answer me? Is it not, then, of the utmost importance to all the gentlemen involved? (Kierkegaard 2000: 112)

Kierkegaard is not right about much, but his first instinct after he finds himself in existence—to register a complaint or two—is certainly justified (see, e.g., genocide, cancer, tsunamis, pedophilia, Alabama, etc.). After all, no one has ever asked to be brought into existence, and one might hope that a situation in which one is compelled to be involved would be better, on the whole, than what has so far transpired on this planet.[1]

However, to complain is to do something. If you are going to complain, is that the first thing you should do? Should you complain at all? This train of thought should inevitably lead to the more general question: What should I do? After a bit of reflection on this question, a slightly more specific question that Kierkegaard hints at should come to mind: What should I do in order to have

[1] Friedrich Nietzsche (1966: 344) refers to it as "that gruesome dominion of nonsense and accident that has so far been called 'history.'"

my life go well for me? Or, in more general terms: What makes a life go well for the person who lives it? I think this is *the* most important question. That is a bold claim. Why, you might ask, do I think this is the most important question? I will get to that answer in due course, but first let us get a better understanding of the question.

Some lives go better for the people who lives them than other lives. This is an idea that is accepted by everyone, or quite nearly everyone, until it is brought up in a philosophical context. A simple thought experiment should help to make this clear. Suppose I had come to you before you began to read this and asked you the following question: Would you please tell me how your life could go such that it would be as good as possible of a life for you, or, alternatively, such that it would be as bad as possible for you? You might have little to say at first, but this would be due to your being caught off guard and wanting more time to think about your answer (much like you would be if a genie popped up and asked what your three wishes were), but it would not be because you thought there was no answer to this question.

One mistake you may make in answering this question, at least from the philosopher's point of view, is that you may mention something that will make your life go better, but only contingently so. For example, you may identify having more money as something that will make your life go better. If this is true for you, it is merely contingently true. In other words, having more money might make your life go better, but only in light of your particular circumstances at the time. For instance, if, at the same time I deliver your

additional funds, all minds other than yours go out of existence, then the money will not make your life go better, as you will have no use for money *qua* money. Another way of saying this is that money is only instrumentally valuable (i.e., it is valuable only for what it will lead to). What the philosopher wants to know about this question—commonly referred to as the question of personal welfare, well-being, or prudential value—is what is intrinsically valuable for a person to have (i.e., what is valuable in and of itself for a person independently of any consequences or effects it may have).

One might object to this way of conceptualizing the problem at the outset by claiming that a list of instrumentally good things, like money, would answer the relevant question just as well since these things do lead to the intrinsically good things (otherwise, obviously, they would not be instrumentally good). The first problem with this approach is that these instrumental goods will be very different for different people. For example, giving $10,000 to almost any graduate student I know will be instrumentally good for them, but it will almost certainly not be instrumentally good for Bill Gates. It is hard to tell a story about how an additional $10,000 will lead to anything at all of value for him. This indicates a more basic problem with this approach. Money, to the extent it is instrumentally valuable, may be classified as such only because it can lead to the attainment of things that are intrinsically valuable. In other words, instrumental value can be explained as such only in relation to intrinsic value. Intrinsic value does not need instrumental value to be explained and, accordingly, intrinsic value is the more basic concept.

Therefore, my project will be to locate the thing or things that are intrinsically valuable for a person such that having it or them makes a person's life go better for that person. This is, of course, the task for any theory of personal welfare if it is to be a theory of welfare at all. But why do we want a theory of personal welfare at all? After all, we could agree that some lives go better than others, but still maintain that the theory that accounts for this fact is unimportant. There are at least three related reasons why I think this question—what makes a life go best for the one who lives it—is the most important question one can ask and, accordingly, why everyone should care about what the correct theory of personal welfare is.[2]

First, suppose someone were to stop you on the street and ask, "Do you want to have a good life?" Your first inclination would probably be to think that someone was attempting to have a laugh at your expense. After all, why would anyone stop you and ask you a question with such an absurdly obvious one-word answer? However, the next question, after you give the almost obligatory answer of "yes" to the first, should make the practical importance of this line of questioning clear: "What, then, makes a good life for the one who lives it?" Here, it seems to me, everyone should grasp the critical significance of this question as it relates to his or her own life. Indeed, how is one supposed to go about having a good life if one has no idea what having a good

[2] Later I will claim that there cannot be valid claims concerning what you *should* care about unless something is known about what you *do* care about. Since you are reading this, I will assume that you care about, at a minimum, a life—either your own or someone else's.

life consists of? Hitting a target without knowing where the target is or what it looks like is purely a matter of luck. And while luck may (or may not) play a role in living a good life, it should not be the star in a one-man show.

Another reason to be concerned with personal welfare involves the lives of the people closest to you, the lives of people whom you will never know, and everyone in between; for you will probably find yourself in a position from time to time of wanting to do something that will make the lives of these people go better for them. You may even, at times, want to do something that will make their lives go worse (e.g., you may want to punish a child or someone who has wronged you).[3] Also, the people in your life will probably ask you for advice on occasion regarding what they consider to be very important issues in their lives. How should you go about these tasks? What is the essence of making a life go better or worse? How will you know if you succeeded or failed in your attempt?

Finally, personal welfare is an essential element of any plausible moral theory. Supposing one wants to be a morally good person, how is one to go about this task? What makes morally right acts morally right? Here one might want a fully general moral theory to guide one's actions. If so, one aspect of evaluating a moral theory will probably be how the welfare of persons at least, and maybe that of all sentient beings, figures into the theory. To see why

[3] It is, of course, very controversial whether it is ever morally permissible to try to make someone's life go worse. This clearly should not be the goal with punishing a child. The case is less clear and more controversial when it comes to punishment of adults guilty of a criminal offense. (See David Boonin's *The Problem of Punishment* for a good discussion of whether it is ever morally permissible to attempt to make someone's life go worse.)

personal welfare needs to be taken into account in an adequate moral theory, consider the following fully general moral justification: WP—An act is morally right if and only if it produces a wooden pickle. WP is absurd, and many objections could be leveled against it. However, one of the main objections surely must be that it gives no consideration to the welfare of any person. This objection seems fatal to any moral theory against which the objection may be leveled.[4]

Now that we have a better idea of what a theory of personal welfare is, we should take a brief look at what it is not. A theory of personal welfare is not a theory about morality generally, nor is it a theory about what makes the world a better place. So there is nothing incoherent, on its face, about the statement that action X made Xander's life go better for him, even though it was a morally wrong act or it made the world a worse place, or both. However, it may turn out that a theory of personal welfare will be connected with these other concepts. For example, one could make the claim that being good for the world is solely a function of personal welfare—a theory referred to as "welfarism." In the same way, one could claim either that personal welfare is a function of personal moral goodness—a view held by Aristotle, among others—or that personal moral goodness is a function of personal welfare—a view held by Ayn Rand, among others. The point here is that all these ideas require claims in addition to those about personal welfare and, accordingly, are distinct from

[4] Amartya Sen (1985: 185) makes a similar point when he says, "It would, of course, be altogether amazing if moral goodness had nothing to do with well-being."

claims about personal welfare. A theory of prudential value is a theory about what has intrinsic value *for* the particular person under consideration. A little imagination should be able to produce scenarios in which other good things, if there are any, are not good *for* any particular person.

Before we proceed down the tracks, there are a couple of objections to this project as a whole that should be addressed. One is the epistemic objection (i.e., how do you know if one life goes better than another?), and the other is the metaphysical objection (i.e., there is no fact of the matter of whether one life goes better than another). As to the latter objection, I have a difficult time believing anyone actually believes this. In order to maintain this objection, you must also maintain that it is not true (or false) that Albert Einstein had a better life than a child born in Poland in the 1930s, herded into the Warsaw Ghetto, and then gassed in the early 1940s in Auschwitz. I could continue with examples, but if that one leaves you cold then it is not clear you can be warmed. Whatever the source of the anxiety over this question, the metaphysical objection is not the most plausible place to exit the train.

The epistemic objection, on the other hand, is more promising. After all, it is always an interesting and valid question for all knowledge claims: How do you *know*? There are a couple of comments to make here. First, the point of the objection cannot be either that questions such as these should not be pursued or that pursuing them has no value on the basis of the answers being dubitable. If that were the point of the objection, then that same objection would be equally devastating for *every* interesting question about our lives and

the world we inhabit. After Descartes's (1996: 15) Evil Demon, even the things we take most for granted about ourselves—other than, perhaps, our own existence—are dubitable. So the fact that all the answers to a question have some level of doubt attached to them cannot serve as a reason to reject the question itself unless one wants to alternate between pondering the mere fact of his own existence and working out math problems. David Hume (1993: 73) provides the right answer on this point: "A wise man . . . proportions his belief to the evidence." And evidence, mostly in the form of arguments, is what you will find in the pages that follow.

The other point to make about the epistemic objection is made quite well by Bertrand Russell (1988: 156-61):

> The value in philosophy is, in fact, to be sought largely in its very uncertainty. The man who has no tincture of philosophy goes through life imprisoned in the prejudices derived from common sense, from the habitual beliefs of his age or his nation, and from convictions which have grown up in his mind without the co-operation or consent of his deliberate reason.
>
> . . . .
>
> Philosophy is to be studied, not for the sake of any definite answers to its questions, since no definite answers can, as a rule, be known[5] to be true, but rather for the sake of the questions themselves; because these questions enlarge our conception of what is possible, enrich our intellectual imagination and diminish the dogmatic assurance which closes the mind against speculation.

[5] Russell should have claimed that we cannot be *certain* about the answers, rather than claiming that we cannot *know* the answers. Although there is a big difference between these claims, his point is still well taken.

I suppose the bottom line for me is that, regardless of the inherent difficulties, the issue of personal welfare is one that everyone should be interested in, particularly once a person understands how intimately it touches her life.[6]

[6] And to those people who are not interested, I propose starting a new era of name-calling (philosophy is woefully deficient in the fun that is *ad hominem* attacks). Aristotle (*Metaphysics*: 4.1006a) called people who accept contradictions "vegetables"; I propose calling the people who are not interested in the question at hand "fruits." Having these kinds of fruits and vegetables in your life—while perhaps intermittently entertaining—will probably make your life go worse.

CHAPTER ONE: OBJECTIVE-LIST THEORY & HEDONISM

So what does make a life go better for the person who lives it? Hopefully, both the import of the question and the question itself are now clear. And so the search begins. Of course, how one goes about looking for something will usually have some effect on how successful one is in finding it. The methodology I intend to follow will be to first set out and evaluate some of the established theories of personal welfare. In so doing, I will consider various objections to the theories, but will focus on those that give the most insight into and advance us toward an adequate theory. During this process, I hope to propose and then refine some basic and essential personal welfare principles that any adequate theory will have to properly countenance. Finally, I will put forward my own theory, extol some of its virtues, and defend it from various objections.

I. OBJECTIVE-LIST THEORIES

When discussing the lives of others and assessing how well those lives may be going, we often make assumptions about what things are good for a person to get. One of the most common examples in twenty-first century America is going to college. So we will often hear that a parent worries about her child going to college, hopes her child will go to college, expects her child to go to college, etc. The same sorts of sentiments can be heard expressed about, to varying degrees, getting married, having children, accepting a certain religion, getting a certain type of job, buying a house, having a certain sexual

orientation, etc. Now of course, the justifications for such ideas will vary from person to person, but I want to focus on just one of these justifications for the time being; specifically, the idea that these things—college, kids, marriage—are intrinsically good for a person to have. In other words, take a person, add these things to her life, and—Shazam!—her life goes better.

This idea has a long intellectual heritage. Philosophers have been putting together such lists for thousand of years, although, interestingly, these lists usually fail to include college, kids, or marriage. In order to give this type of theory, called an *objective-list theory*, its due and to see where it might have difficulties, several examples of these theories are set out below:[1]

In a late Socratic dialogue, Plato (*Philebus*: 66a-66c) set out five classes of intrinsic goods in descending order of goodness:

1) "Measure, and the mean, and the suitable, and the like";
2) "The symmetrical and beautiful and perfect or sufficient, and all which are of that family";
3) "Mind and wisdom";
4) "Sciences and arts and true opinions"; and
5) "Pleasures which were defined by us as painless, being the pure pleasures of the soul herself, as we termed them, which accompany, some the sciences, and some the senses."

G.E. Moore, in *Principia Ethica* (1903: § 113) had this to say about intrinsic value:

> By far the most valuable things, which we know or can imagine, are certain states of consciousness, which may be roughly described as the pleasures of human intercourse and the enjoyment of beautiful objects. No one, probably, who has asked himself the question, has ever doubted that personal affection and the appreciation of what is beautiful in Art or Nature, are good in

[1] For an example of a defense of objective-list theories generally where no list is actually specified, see Richard Arneson (1999: 113-42).

> themselves; nor, if we consider strictly what things are worth having *purely for their own sakes*, does it appear probable that any one will think that anything else has nearly so great a value as the things which are included under these two heads. . . . That they are truths—that personal affections and aesthetic enjoyments include all the greatest, and by far the greatest, goods we can imagine, will, I hope, appear more plainly in the course of that analysis of them, to which I shall now proceed.

In *The Right and the Good*, W.D. Ross (1930: 102) writes:

> Four things, then, seem to be intrinsically good – virtue, pleasure, the allocation of pleasure to the virtuous, and knowledge (and in a less degree right opinion). And I am unable to discover anything that is intrinsically good, which is not either one of these or a combination of two or more of them.

Martha Nussbaum, in *Sex & Social Justice* (1999: 41-42), "claims that a life that lacks any one of these capabilities, no matter what else it has, will fall short of being a good human life":

1) Life;
2) Bodily health and integrity;
3) Bodily integrity;
4) Senses, imagination, thought;
5) Emotions;
6) Practical reason;
7) Affiliation;
8) Other species;
9) Play; and
10) Control over one's political and material environment.

Finally, Derek Parfit (1984: 499) sets out, but does not endorse, an example list and describes the theory this way in *Reasons and Persons*:

> Turn now to the third kind of theory that I mentioned: the Objective List Theory. According to this theory, certain things are good or bad for people, whether or not these people would want to have the good things, or to avoid the bad things. The good things might include moral goodness, rational activity, the development of one's abilities, having children and being a good parent, knowledge, and the awareness of true beauty. The bad things might include being betrayed, manipulated, slandered, deceived,

> being deprived of liberty or dignity, and enjoying either sadistic pleasure, or aesthetic pleasure in what is in fact ugly.

There are several objections that could be made against these particular lists, or against the notion of a list in general, that I do not intend to pursue at length here for the reason that these objections will not advance our search to any significant degree. I will call these objections the objections of *authority*, *commensurability*, and *completeness*. The authority objection will likely be the first one almost any class of undergraduate ethics students will come up with: Who gets to decide what goes on the list and what does not? To bring this area into line with other fields of study, the question is probably more properly framed as *why* are these items on the list? Is it just a brute fact, a fact not explicable in terms of anything else, that these items are on the list or is there some common theme that links some or all of the items together? The commensurability objection involves the apparent difficulty in comparing the lives of the people that have these goods, and in comparing the value of the various goods within a single life. Consider Nussbaum's list above. Let us suppose that Willie gets five of the things from the list on a consistent basis for 60 years and that Marcus gets all ten of the things on the list for 45 years. Whose life went better for him? Are some things on the list better than others (e.g., bodily integrity vs. play) such that a small amount of one outweighs a large amount of the other? Do we need to know which five Willie got in order to know whose life went better? There are other questions of this sort that I will not pursue here. Finally, there is the objection of completeness. For an objective-list theory—or any other theory—to be a complete theory of personal

welfare (i.e., a theory that tells us how well any life will go, is going, or has gone for the person who will live, is living, or has lived it), the list of intrinsic goods for a person must be complete. So if Nussbaum is right and other species partly determine human welfare, then if there are two identical lives with the only difference being that one had a puppy for a day, then the person-with-the-puppy life will have gone better. This will be a fatal objection to the completeness of any of the other theories listed above.

Leaving these objections to one side along with the questions about how devastating any or all of them are, the objection I intend to pursue at length is one hinted at by Parfit's quote above: "According to [objective-list] theory, certain things are good or bad for people, *whether or not these people would want to have the good things, or to avoid the bad things*" (Parfit 1984: 499) (emphasis added). Now, to be sure, there will be a very large number of people who will want most or all of the items on these lists. If people did not want these things, these theories would probably need to be accompanied by a fairly robust error theory that would explain why so few people wanted the things that were intrinsically good for them to have. The important thing to notice about objective-list theories, as Parfit rightly points out, is that people's attitudes toward the things on the list are *entirely* irrelevant. This is a necessary consequence of all of these theories, as they have not left themselves any conceptual space or framework to handle actual attitudes. An objective-list theorist might respond to this by saying, "Actual attitudes do not matter. The things on the list just are intrinsically good for a person to have.

Therefore, a person who has properly functioning faculties[2] *will* want the items on the list."

Putting aside the question of whether a person who does not care about the things on the list is malfunctioning in some way, this argument still fails. As Harry Frankfurt (1999: 158) points out:

> A person who acknowledges that something has considerable intrinsic value does not thereby commit himself to caring about it. Perhaps he commits himself to recognize that it *qualifies* to be *desired for its own value* and to be *pursued as a final end*. But this is far from meaning that he does actually desire it or seek it, or that he ought to do either. Despite his recognition of its value, it may just not appeal to him; and even if it does appeal to him, he may have good reason for neither wanting it nor pursuing it. Each of us can surely identify a considerable number of things that we think would be worth doing or worth having for their own sakes, but to which we ourselves are not especially drawn and at which we quite reasonably prefer not to aim.

Whether we are inclined to agree with Frankfurt or not, there are certainly going to be people who lack any sort of positive mental attitude toward the items on the objective list in question. The objective-list theorist has a dilemma with regard to people in this class. The first option is to claim that these people's lives go better when they get the things on the list no matter how they may feel about it. If an objective-list approach is to be retained, then this will have to be the option chosen. The second option is to admit that a person's life does not go better by getting something regardless of what attitude she takes toward it. This, I think, has to be the right answer, for a reason already discussed. Recall that we are looking for value *for a subject*. It seems

[2] Some caveat of this sort will be required. This may not be what an actual objective-list theorist would want to claim, but this will not matter, as all caveats of this type will suffer from the same infirmity.

the objective-list theorists lose sight of this point. Perhaps what they really have in mind is a theory about what makes the *world* a better place. Maybe the world is a better place if people get the things on the objective list and an even better place if they take the appropriate amount of pleasure in getting just those things.[3] While this revision of the theory may be plausible for the value of possible worlds, the actual objective-list theory of personal welfare fails because it does not properly account for the fact that we are looking for how well a life went *for the person who lived it*. The problem here is not simply that a person might have no discernible attitude toward getting the items on the list, although this in itself could not be counted as a virtue for a theory of prudential value. The problem is that a person may positively detest any or all of the things on the list. If so, is it really plausible to claim that such a person's life is going better *for him* if he gets those things? No. And the intuition behind this answer will provide us with the most essential foundational principle that any adequate theory will have to heed.[4]

A great deal of care should be taken in formulating any principles that will serve as an essential requirement for an adequate theory. Another way of thinking about this is that these principles are a way of defining the concept of

---

[3] Parfit (1984: 502) theorizes that this is what makes a life go best, but his comments could be easily altered to reflect this idea.

[4] Robert Merrihew Adams (1999: 95) sums up the guiding intuition here nicely:

> Another truth about human well-being that is intuitively evident is that a person's good is not very fully realized unless she likes or enjoys her life in the long run. You may be very virtuous; you may be brilliant, beautiful, successful, rich, and famous; but if you do not enjoy your life, it cannot plausibly be called a good life *for you*.

personal welfare. So, yes, a theory of personal welfare is a theory about what makes a life go best for the person who lives it—but what does *that* mean? Here I should like to steal David Boonin's requirements for a definition wholesale (Boonin 2008: 4-5). First, and almost too obviously, we want a definition to be *accurate*. The definition should supply the necessary and sufficient conditions for cases of personal welfare, should be reasonably in line with ordinary uses of these words, and should help us solve the borderline cases where it is unclear if someone's life is going better or not. Second, a definition should be *illuminating*. In other words, we want a definition to point us in the right direction in terms of getting a better understanding of the concept and any necessary implications of the concept as defined. Finally, we want a definition to be *neutral*. In other words, we do not want to *unfairly*[5] beg

---

[5] Michael Huemer (2005: 69-70) nicely illustrates when "begging the question" is not actually begging the question—and thus is fair—in the following excerpt:

> It is not the case that whenever an argument deploys a premise that directly and obviously contradicts an opponent's position, the argument begs the question. Still less is it true that whenever a consistent opponent would reject at least one of an argument's premises, the argument begs the question. (The latter condition applies to every valid argument.) Consider another famous philosophical argument: Philosopher *A* claims that 'knowledge' means 'justified, true belief'. Philosopher *G* points out the following sort of example:
>
> > Suppose that Smith justifiably believes that Jones owns a Ford (he has often seen Jones driving a Ford, has seen the title to the car, and so forth). Smith correctly infers from this that the following statement is also true: 'Jones owns a Ford, or Brown is in Barcelona'. (Barcelona was selected randomly; Smith has no idea where Brown is.) But suppose that, improbably enough, Jones actually does *not* own a Ford; it was sold just a few minutes ago. But by pure coincidence, Brown happens to be in Barcelona. In this

the question when coming up with our definition, or in this case, with our principles. This is particularly critical with regard to the first principle of personal welfare contained in the next paragraph. To illustrate the critical nature of this requirement, suppose we claimed that any acceptable theory of personal welfare had to be objective. Should we then be surprised to find that our search ended with an objective-list theory being the right one? No, but this would be a case of unfairly begging the question, as the objective requirement seems to lack any independent motivation.

So what is the most accurate, illuminating, and neutral principle of personal welfare? As we saw with our analysis of objective-list theories, the right account of personal welfare must adequately countenance the actual

---

> case, Smith believes [Jones owns a Ford, or Brown is in Barcelona]. But intuitively, Smith does not *know* [Jones owns a Ford or Brown is in Barcelona]. Thus, justified, true belief is not the same as knowledge.

The above argument is widely, and rightly, taken to conclusively refute the 'justified, true belief' account of knowledge; indeed, it is one of the few, and one of the most celebrated, examples of a conclusive refutation of a previously widely-held view in modern philosophy. How weak it would be for *A* to reply:

> *G* has merely begged the question. I say that knowledge is justified, true belief. From this, it directly and obviously follows that Smith *does* know [Jones owns a Ford or Brown is in Barcelona] in *G*'s example. For *G* to assert that Smith lacks such knowledge just assumes that my definition is wrong. So *G* has proven nothing.

A is correct to note that *G*'s premise directly and obviously contradicts *A*'s theory. But this does not mean that *G* begged the question; it means only that *G*'s refutation of *A* was direct and obvious. *G* succeeds while *A* fails because *G*'s premise, once stated, is intuitively obvious, while *A*'s theory is not intuitively obvious but needs to be tested by considering examples. *A* thus is not in a position to simply appeal to his theory as a justification for denying the premise that would be used to refute the theory.

mental states of the person under consideration. But what kind or kinds of mental states should be taken into account? Although "want," "pleasure," and "desire" were the most commonly used terms in previous paragraphs, the continued use of these terms seems unduly restrictive, such that we may be unfairly begging the question by continuing to use them. Accordingly, I propose that we begin with a stance that is completely neutral regarding which mental state or states may ultimately be included in any adequate theory:

> INTERNALIST PRINCIPLE (IP) – The intrinsic value of a life (or part of a life) for the one who lives it is at least partly determined by the actual mental states of the person living that life (or life-part).

IP is meant to be ambiguous, as I have only attempted to take the most tenuous step in the right direction in formulating this principle. IP merely makes the claim that the mental states of X are relevant, at least to some degree, to the question of how well life goes for X. This will allow two distinct and competing camps of welfare theories that rely on mental states to stay in the running.[6] The first is the hedonist camp, which will include states like pleasure, enjoyment, satisfaction, and the like. The other camp I would only characterize as, borrowing a term from Donald Davidson (2001: 4), the pro-attitude camp. The mental states in this camp will include desires, urges, values, goals, moral views, and the like. The intent with IP is to be as inclusive

---

[6] IP, in its current form, does not even necessarily rule out some of the objective-list theories above. For example, Ross's (1930: 102) theory might satisfy IP's mental state requirement quite well. However, all objective-list theories will be eliminated in the coming pages as we scrutinize, and ultimately revise, IP.

as possible with respect to any mental state that could be, or that could be part of, a state of affairs that is intrinsically valuable for a person.

II. HEDONISM

Before explicitly setting out the theory of hedonism, it is worth noting one caveat concerning what hedonism is *not*. Any connotations associated with the common usage of the term should be set aside. To be sure, the concept of hedonism as defined by common usage and as defined by philosophers had a common origin, but the conceptual space that these two terms now occupy has little to no overlap. Hedonism, at least in this context, has nothing to do with occasionally clothed, mind-altered funsters on a beach in some exotic, libertine paradise. With that in mind, let us turn to an examination of what hedonism is.

The seeds for modern hedonism can be found in the works of the great utilitarians Jeremy Bentham and John Stuart Mill. Part of the argument for their ethical theory—and the largest part of the foundation—consists of claims about the value of pleasure and pain. According to Mill (2006: 335): "The utilitarian doctrine is that happiness is desirable, and the only thing desirable, as an end; all other things being only desirable as a means to that end." Mill (2006: 320) then goes on to explain what he means by happiness: "By happiness is intended pleasure and the absence of pain; by unhappiness, pain and the privation of pleasure." Mill (2006: 320) also combines these ideas into this claim: "pleasure and freedom from pain are the only things desirable as ends; and that all desirable things . . . are desirable either for the pleasure

inherent in themselves or as a means to the promotion of pleasure and the prevention of pain."

However, these claims do not get us all the way to the form of hedonism we are interested in here. Recall that in axiology we are looking for what thing or things are intrinsically good for a subject. Mill has only made claims about what is intrinsically *desired* by a subject. Accordingly, in order to get from the claim that only pleasure is intrinsically desired by a subject to the conclusion that only pleasure is intrinsically good for a subject, Mill needs an additional premise equating being intrinsically desired with being intrinsically good. And this is clearly what Mill has in mind, as he often implicitly equates well-being and happiness. In other words, the good life is the happy life and the happy life is the more pleasurable life.

In the modern era, this idea is best illustrated by what Fred Feldman, a hedonist, calls Default Hedonism (DH). Feldman puts forward DH in his book *Pleasure and the Good Life* as a starting point because he thinks most formulations of hedonism in the professional literature are defective in one way or another.[7] An adequate formulation of hedonism, DH is set out as follows:

i. Every episode of pleasure is intrinsically good; every episode of pain is intrinsically bad.

ii. The intrinsic value of an episode of pleasure is equal to the number of hedons of pleasure contained in that episode; the intrinsic value of an episode of pain is equal to – (the number of dolors of pain contained in that episode).

[7] As an example of a defective formulation of hedonism, Feldman (2004: 22-25) quotes the version in William Frankena's *Ethics* (1973: 84) and then recites six ways that it fails to be an adequate characterization.

> iii. The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of pleasure and pain contained in that life, in such a way that one life is intrinsically better than another if and only if the net amount of pleasure in the one is greater than the net amount of pleasure in another. (Feldman 2004: 27)

Hedonism, at least at first blush, seems to have several points in its favor. First and foremost, there is a great deal of intuitive plausibility to the idea that pleasure makes a life go better and pain makes a life go worse. Illustrations and calm reflection should not be required to see this point. Second, the theory is simple and, therefore, easy to understand. As shown above, the theory can be set out in a few sentences, which makes it easy for a beginner to be up and running with the basic ideas of hedonism in a very short time. Third, there are no glaring[8] commensurability problems with hedonism as there are with, using an earlier example, objective-list theories. Finally, DH makes it very easy, at least in theory, to calculate how well a life (or any part thereof) is going for the one who lives it. A list of a person's hedons and dolors and a calculator is all you need.

The simplicity of hedonism, while one of its greatest virtues, is also cited as leading to the theory's greatest defects. The challenge, roughly stated, is to

---

[8] In order to avoid problems of commensurability, hedonists will often describe pleasure and pain as being "opposites" (see, e.g., Feldman 2004: 26). This is supposed to make it acceptable to directly compare units of pleasure to units of pain to come up with a net balance of pleasure vs. pain. However, it is not clear that, say, some pains can be compared to other pains in this way, much less that pleasures can be compared to pain in this manner. So, for example, can the pain of breaking an arm be compared to the pain of losing a loved one? Or can the pleasure of an ice cream cone be compared to the pain of stubbing a toe? These are questions that I will have to leave unexplored here.

question whether prudential value is really so simple after all. This can be summed up in two related questions that I intend to pursue next. First, does a person's well-being depend, at least to some degree, on something other than that person's mental states? Second, does a person's well-being depend, at least to some degree, on mental states other than pleasure and pain?

DH claims the only things that are relevant to determining personal welfare are certain mental states. One way to evaluate the plausibility of this aspect of hedonism is to create a scenario where, in essence, mental states are the only things a person has. Robert Nozick (1974: 42) did just that with his "experience machine":

> Suppose there were an experience machine that would give you any experience you desired. Super duper neuropsychologists could stimulate your brain so that you would think and feel you were writing a great novel, or making a friend, or reading an interesting book. All the time you would be floating in a tank with electrodes attached to your brain.

The rest of the story rounds out the thought experiment quite nicely from a phenomenological perspective—you can pick your experiences from a huge library of desirable experiences, you will not know you are plugged into the machine while you are plugged in, your loved ones can plug in as well so there is no reason to stay unplugged in order to serve them, etc. Nozick then asks, "Would you plug into the machine?"

If you accept hedonism as the correct theory of personal welfare and you want your life to go better than it otherwise might, then you should plug in. Or is this conclusion a bit too hasty? After all, the only things that matter from a

hedonistic standpoint are certain mental states, and an experience machine is not necessary for mental states. At this point there is no reason to plug in, just as there is no reason *not* to plug in. But it is not just the *existence* of mental states that matters. According to hedonism, it is the type, duration, and intensity of the mental states that make your life go better. And this is just exactly what, presumably, you can be guaranteed a better mix of in the experience machine. If pleasure and the avoidance of pain are what determine the intrinsic value of your life, it seems almost assured that you will be better at picking experiences off a list that will produce these outcomes than at orchestrating your life—and all of the things that affect your life over which you have little to no control—to produce a more favorable ratio of pleasure to pain.

Nozick (1974: 43), for one, would not plug in for the following reasons:

> What does matter to us in addition to our own experience? First, we want to *do* certain things, and not just have the experience of doing them. . . . A second reason for not plugging in is that we want to *be* a certain way, to be a certain sort of person. Someone floating in a tank is an indeterminate blob. There is no answer to the question of what a person is like who has long been in the tank. Is he courageous, kind, intelligent, worthy, loving? It's not merely that it's difficult to tell; there's no way he is. Plugging into the machine is a kind of suicide.

So Nozick's reasons for not plugging in are that you cannot do certain things and that you cannot be a certain way. Yet consider the following scenario from Shelly Kagan (1998: 34-35):

> Imagine a man who dies contented, thinking he has achieved everything he wanted in life: his wife and family love him, he is a respected member of the community, and he has founded a successful business. Or so he thinks. In reality, however, he has been completely deceived: his wife cheated on him, his daughter

> and son were only nice to him so that they would be able to borrow the car, the other members of the community only pretended to respect him for the sake of the charitable contributions he sometimes made, and his business partner has been embezzling funds from the company, which will soon go bankrupt.

Here again we have, presumably, a very good life if the only things that matter for well-being are mental states. Notice that, *in terms of how it feels from the inside*, this man's life is *exactly* the same as that of a man who actually has the things that our subject merely believes he has. If the only thing that matters is how a life feels from the inside (i.e., mental states), then there would be no reason to prefer the life of the man who actually has these things to the life of the man who just believes he does. Furthermore, we have resolved Nozick's issues concerning actively doing certain things and actively being a certain way. Is it in fact the case that the lives of these two men go equally well? If we were one of these two men, should we be indifferent as to which one of them we were? A hedonist might prefer the life of the man who actually has the things that he thinks he does simply as a result of having to choose one, even though there is no reason to prefer one to the other; however, he will have to prefer the "false" life if we change the "true" life to include just one feature that seems worse from the inside (e.g., one more stubbed toe or mild headache). From these hypotheticals one could reasonably conclude that features beyond our conscious experience can have an impact on how well our lives go.[9]

But hedonism's failure to give any weight at all to the outside world may not be the only, or even the best, reason to call it into question. Recall that,

[9] I will have much more to say on this topic in Chapter Four.

according to hedonism, not all mental states count for personal welfare. It is only one range of mental states, the continuum of pleasure to pain, that makes one's life go better or worse. To evaluate this aspect of hedonism, consider the following:

> Maxwell Edison, who majored in medicine in the late 1960s, has always had a very strong affinity for Paul McCartney. Maxwell, a committed (and very misunderstood) hedonist, has devoted his medical career to finding a way to make people's lives, and Paul's in particular, go better. After a few failed attempts involving Joan, his teacher, and a judge, Maxwell has perfected his technique. He sneaks up behind Paul and strikes him with a silver hammer *just so*. While this blow reduces Paul's cognitive abilities to that of an infant, it simultaneously extends his life span by 10 years and causes him to experience the most intense pleasure he has ever experienced over and over for the rest of his days. Paul, a billionaire, will be cared for extraordinarily well for the remainder of his life.

If hedonism is the correct theory, then Maxwell has very clearly succeeded in making Paul's life go better. Assuming your needs would be met in a way similar to Paul's, would you choose to be subjected to Mawell's now perfected Silver Hammer Therapy™? I suspect that no one reading this thinks Maxwell has improved Paul's life and, therefore, no one would choose to be so struck about the head. The most likely reason for the reticence to undergo this treatment is that you value things, even if they are only other mental states, in addition to and perhaps more than unending and unwavering pleasure.

What the experience machine, the businessman, and Mawell's Silver Hammer Therapy™ suggest is that perhaps something other than, or at least in addition to, pleasure and pain affects how well a life goes for the one who lives it. Here we have three options for how to integrate these ideas into our theory

of personal welfare. The first option is to ignore or explain away these ideas, ultimately concluding that, in fact, pleasure and pain are the only things that determine prudential value. A hedonist will have to give this response if he is to remain a hedonist. For those people who found one or more of our experience machine line of cases compelling, this response will be deeply unsatisfying. This is because the hedonist is saying, in essence, "You who think something other than pleasure and pain is relevant to your personal welfare are wrong. I, the hedonist, know better than you what will make your life go better." In this way, hedonism is like an objective-list theory. The main differences are that hedonism's list is much shorter than that of the typical objective-list theory, and the problem just described is much more obvious when it comes to objective-list theories. Recall that our Internalist Principle (IP) requires an adequate theory of personal welfare to take into account the actual mental states of the person living that life. IP was designed to be broad enough to allow the pleasures and pains of the hedonism camp to count. What our experience machine line of cases suggests, then, is that IP is too broad since it is still subject to the same sort of objection that prompted the adoption of IP in the first place.

To see how IP should be narrowed to accommodate these ideas, it is important to be clear about the differences, if any, between the objection to objective-list theories and the objection to hedonism. So suppose there were an objective-list theory that included just three items: sex, drugs, and rock & roll. We then find a person whose life was filled with sex, drugs, and rock &

roll who states that he hated just about every moment of his life. Based on this report, it seems wrong to conclude that his life did in fact go well for him. It would be odd to say to him, "You are wrong about your life. It did go well for you. We here at the Sex, Drugs, and Rock & Roll Institute—after many nights of research—are sure of it." The odd tenor of this remark stems from its very strong paternalistic character. If IP is the most basic principle that an adequate theory of personal welfare must meet, the paternalism objection is the most basic objection that an adequate theory must avoid. This should not be surprising since they are flip sides of the same coin. A theory of personal welfare, in order to be a theory of personal welfare, must identify some thing or things that make a life go better for the one who lives it. An adequate theory should always be mindful of trying to reduce the size of the conceptual gap between what the theory identifies as making a life go better and what the person who is living the life cares about. For a gap too large, the paternalistic response sounds tone deaf, if not completely bizarre.

For example, suppose there were an objective-list theory that included just one thing: Merchant Ivory films. Insisting that someone who watched *The Remains of the Day* every day of her life and despised it more with each viewing had a great life probably makes one a good candidate for a room with padded walls. However, the paternalistic response is much more plausible when it comes to smaller gaps. This is the case with hedonism and is why the paternalism problem is much less obvious. Whereas objective-list theories typically completely disregard how anyone feels about the items on the list,

hedonism's list—pleasure and the absence of pain—ensures that at least the good life for the one who lives it will be the pleasurable life. And while the conceptual gap is smaller here, there is a gap nonetheless. So suppose there is a woman who finds sex to be a very pleasurable activity, but does not care at all about having sex and would always prefer a good nap instead. Here again it sounds odd, albeit less so, to insist to her that her life is going better for her if she has sex instead of taking a nap, even though this is contrary to her wishes. We can further suppose that her preferences are similar with regard to all other sensory pleasures; she prefers work to a good meal, study to a massage, etc. Is it plausible to claim that all of her preferences make it the case that her life is going worse when she gets what she wants?

The paternalism objection also lurks for the second of these potential responses to the experience machine line of cases. This response is to claim that just this list of things, other than pleasure and pain, necessarily impacts how well a life is going for the one who lives it. Depending on what is on this list of other things that are supposed to affect prudential value, a conceptual gap may open between what the person cares about and what is supposed to determine that person's welfare. The paternalistic response will be plausible, perhaps, for some of these things and less so for others.

It is the third response to the experience machine line of cases that holds the most promise for directing us how to appropriately narrow IP. The third response is to claim that things other than pleasure and pain can matter for

purposes of personal welfare. Given our paternalism concerns,[10] the most obvious answer for when they *can* matter is when the person cares about the thing in question. Notice that the claim here is merely that these things *can* matter if the person cares about them. There will, I think, still be occasions that merit a paternalistic response. However, this does not undermine the fact that a paternalistic answer in axiology is an inherently suspect classification. Given this fact, a general principle concerning paternalism is in order:

> PRINCIPLE CONCERNING PATERNALISM (PCP): Paternalistic claims in axiology must be justified by a compelling theoretical interest and must be narrowly tailored to serve that interest.

A couple of features of PCP are worth noting. First, a compelling theoretical interest is required. This does not require universal agreement among the public at large, much less universal agreement among philosophers. Philosophers cannot even agree that there is an external, mind-independent world, which generally suffices to eliminate any consensus about anything that may happen in the world. Nonetheless, there will have to be a reasonable amount of overlap regarding intuition and a great deal of independent

---

[10] Shelly Kagan (1998: 40) nicely sums up the concern over paternalism as it relates to objective-list theories, stating that such theories provide that

> possession of the objective goods makes one better off—regardless of whether or not one realizes this. This seems to have the implication that your life could be made better off by the possession of some 'good' even though you yourself dislike it and would greatly prefer to be without it: since the good possesses objective value, your own opinion on the subject is quite irrelevant. Your life could be going well even though you are unhappy with almost all its central features!

For additional statements of paternalistic concerns about theories of welfare, see, e.g., Feldman (2004: 17), Sumner (1996: 45), and Railton (2003: 47).

motivation to count as a compelling theoretical interest. Second, the definition of the paternalistic item or items to be included in any personal welfare calculation must be narrowly tailored to include all and only the items that are the subject of the compelling theoretical interest. Any definition found to be either over- or under-inclusive must be refined appropriately or discarded altogether.

With these ideas in mind, we can turn to the task of properly narrowing IP. Recall that IP was intentionally drafted to be broad enough to include mental states that are usually claimed to be the basis for hedonistic theories of welfare along with what Davidson calls "pro-attitude" mental states. This class of mental states, according to Davidson (2001: 4), is composed of the following:

> . . . desires, wantings, urges, promptings, and a great variety of moral views, aesthetic principles, economic prejudices, social conventions, and public and private goals and values in so far as these can be interpreted as attitudes of an agent directed toward actions of a certain kind. The word 'attitude' does yeoman service here, for it must cover not only permanent character traits that show themselves in a lifetime of behaviour, like love of children or a taste for loud company, but also the most passing fancy that prompts a unique action, like a sudden desire to touch a woman's elbow. In general, pro attitudes must not be taken for convictions, however temporary, that every action of a certain kind ought to be performed, is worth performing, or is, all things considered, desirable. On the contrary, a man may all his life have a yen, say, to drink a can of paint, without ever, even at the moment he yields, believing it would be worth doing.

While Davidson is discussing pro attitudes as they relate to action, this list need not be so limited for our purposes. I doubt that one could come up with another pro attitude that could not be subsumed under one of the categories

on this list, but nothing of consequence hinges on this claim, as will soon become clear.

One thing that should be clear by now is that the term "pro attitude" is very broad. This is a good thing at this point, as our mental state or states must surely be on this list. The question now becomes: Do we proceed with the entire list of pro attitudes, or is there a way to eliminate some of the items on the list based on some idea we have already come across? Recall that the problem with hedonism was the conceptual gap between what the person cares about and pleasure and pain. This will clearly be a problem for many of the items on the pro attitude list as well. Take, for example, public values or aesthetic principles. One can easily imagine a person who does not care about these things at all and who claims that every moment spent contemplating these things or acting upon them is extraordinarily unpleasant. Do these things merit a paternalistic response? No. PCP requires a compelling theoretical interest that is lacking here. In other words, there seems to be little to be gained and potentially much to be lost by insisting, in the face of contrary reports from the person at issue, that her life goes better *for her* by getting these things.

So if the problem is going to be the conceptual gap between what a person cares about and whatever item is put forward as necessarily affecting her welfare, then the most successful approach will be one in which the theory proposed does not deviate from what the agent cares about in calculating personal welfare.

There is also powerful independent motivation for this approach. Recall from the prologue that one of the reasons a theory of prudential value is important is that it will tell us how to go about the task of making the lives of those near and dear to us go better for them. We may also be interested in making the lives of some go worse, as in the case of, for example, people guilty of committing a crime. Now suppose there is a person, Veronica, whom you are trying to affect in this way (i.e., you are trying to make her life go better or worse). Without any more information about her, how will you go about affecting her welfare? If you are trying to improve her welfare, the best option is to just give her some money, with the idea being that you do not know what she cares about and money will give her the greatest flexibility in getting some object or experience that she does care about. However, what if you then learn that there is in fact *nothing* that she cares about now and there will in fact be *nothing* that she cares about at any point in the future? How would you go about the task of helping her or harming her then?

The first reaction someone tasked with helping or harming Veronica would have is likely disbelief. This person might think Veronica has had a hard life and was saying these things as a defense mechanism to prevent further harm to herself. So if Veronica were given a billion dollars or were made Ruler of the Universe, she would later admit to secretly wanting these things and caring about them once she received them. Or, more radically, if you soaked her in gasoline and lit her on fire, she would, if she were able, admit to caring about not being in this situation. However, suppose it is

actually the case that Veronica cares about nothing, and the person charged with helping or hurting her personal welfare comes to believe this. The next reaction would probably be to try to coax her into caring about something again. We could imagine this process as approximating the familiar scene in movies where the cop tries to talk the potential jumper off the building or bridge. Accordingly, you might talk to Veronica about her parents, kids, siblings, pets, co-workers, neighbors, other relatives, old hobbies, past favorite foods, books, movies, music, etc., trying to spark some concern in her about one or more of these things. Supposing this effort is unsuccessful, what then? The only reasonable thing to conclude at this point is that it is not possible to make Veronica's life go better or worse. She is, quite literally, beyond help.

The tale of careless Veronica should prove quite useful in restricting IP. Let us suppose that the moral of Veronica's tale produces the following principle:

> PRINCIPLE CONCERNING CARING (PCC): If X does not care about anything and could not be made to care about anything in the future, then it is not possible for X's personal welfare to be affected going forward.

Notice that PCC allows for the fact that X's life may have gone better or worse for X in the past. PCC simply says that at any time X cares about nothing and could not be made to care about anything in the future, there is no way to impact the prudential value of X's life. Given PCC, we can now narrow IP as follows:

> INTERNALIST PRINCIPLE' (IP'): The value of a life (or part of a life) for the one who lives it is determined to a significant degree by what the person in question cares about.[11]

IP' claims that prudential value is determined *to a significant degree* by what the person cares about. IP' does not claim that prudential value is determined *solely* by what the person cares about in order to leave room for a paternalistic response if the conditions for PCP are met. IP' also does not claim that prudential value is determined *to some extent* by what the person cares about because this potentially allows what the person cares about to play too little a role in the theory, thus requiring other paternalistic factors to play a greater role than PCP will allow.[12]

So which of Davidson's pro attitudes does IP' pick out as having the closest possible relationship to caring? The most promising candidate is the first one that Davidson mentions: desire.[13] This is because caring will require desires of some sort, as it is not possible for a person to care about something,

---

[11] IP' is in line with the internalist notions put forward by a couple of recent treatments of this subject. Connie Rosati (1995: 300 n.10) specifically mentions taking into account what the subject cares about for purposes of personal welfare. Similarly, Robert Noggle (1999: 303) claims that welfare calculations must take into account the "agent's own ends and goals."

[12] Of course it is possible, at this point in the proceedings, that paternalistic factors could play a large role in the final theory by satisfying PCP's theoretical requirements. This will turn out not to be the case, as will be shown in the coming chapters by returning to the issue of paternalism on multiple occasions.

[13] The next three items on the list—wantings, urges, promptings—could all be in play here as well, being synonymous with, or very closely related to, desire.

yet have no desires at all. Caring entails desire. We shall examine the question of whether desires entail caring in Chapter Three.

This brings us to the third of the three major theories of personal welfare: desire-satisfaction theory. Desire-satisfaction theories of welfare claim, roughly, that a person's life goes better when the person gets what he wants. We will thoroughly examine different variations of this theory in the next chapter, but for now the preliminary signs seem encouraging. Desire-satisfaction theory seems to allow for the possibility of tracking what we care about (i.e., it fits nicely with IP' and PCC) and does not seem to raise any worrying concerns with paternalism (i.e., it heeds PCP). Yet, as you rightly imagine and as we shall see, desire-satisfaction theories have their own issues.

Before turning to desire-based theories, we should evaluate whether hedonism can be salvaged given the principles—IP', PCP, and PCC—set out in this chapter. Feldman thinks that it can be. After formulating Default Hedonism (DH) as a starting point and noting that "some critics of hedonism seem to understand hedonism in something like this way," Feldman (2004: 27) goes on to say that he is not going to defend DH. Instead, he is going to defend what he calls attitudinal hedonism (AH) (Feldman 2004: 55). The basic difference is that where DH claims that some sensory pleasure is what makes a life go better, AH claims it is attitudinal pleasure that serves this function. "A person takes attitudinal pleasure in some state of affairs if he enjoys it, is pleased about it, is glad that it is happening, is delighted by it" (Feldman 2004: 56). Moreover, according to Feldman (2004: 56), attitudinal pleasures are

always "directed onto objects." It is this last claim, I think, that is supposed to be the main advantage of AH. However, I do not think this feature of AH will ultimately be successful in helping to defend against the types of objections presented above. This is because AH, just like DH, is a form of "mental statism," as Feldman (2004: 67) himself notes. A mental statist theory claims that the prudential value of a life is solely determined by facts about the mental states of the person who lives that life such that if two lives are identical with respect to mental states, then they are necessarily equivalent with respect to prudential value.

While I do think that the same sorts of objections could be pursued against AH as were pursued above against DH, I will not do so here. This is because these objections can be pursued in a more relevant and illuminating fashion against an equivalent desire-satisfaction theory. Chris Heathwood (2006: 559) argues that what he calls the "most plausible form of hedonism," Feldman's Intrinsic Attitudinal Hedonism,[14] is extensionally equivalent to "the most plausible form of desire satisfactionism," Subjective Desire Satisfactionism (SDS). I will accept the claim about the equivalence of IAH and SDS for the purposes of this project. Heathwood's (2006: 559) central claim is that pleasure is the subjective satisfaction of desire. Whatever the truth of this claim, it seem as though Feldman's IAH is an attempt to incorporate some of

[14] This version of AH is limited to intrinsic enjoyment. In other words, enjoyment taken in a state of affairs for its own sake, not enjoyment taken in a state of affairs that merely relates to some other state of affairs the person takes pleasure in.

the central intuitions of desire satisfactionism into hedonism. In so doing, Heathwood noticed that Feldman may have been so successful in this that he turned his "hedonistic" theory into a form of desire-satisfaction theory.[15] And whether IAH and SDS are equivalent or not, they are certainly close to being so. If they are not equivalent, then it would seem to be a charitable interpretation of IAH that would allow this to be the case (although nothing hinges on this claim since the equivalence of the two theories is being accepted). Accordingly, any successful objection against SDS will also be at least as successful against IAH. SDS, along with other versions of desire-satisfaction theories, will be evaluated in the coming pages.

[15] As Feldman (2004: 168-69) himself notes, there is some concern from critics that IAH is no longer a form of hedonism at all.

CHAPTER TWO: DESIRE THEORY

Personal welfare is primarily a function of what we care about. Caring about something entails having desires about it. Accordingly, desire satisfactions will at least be part of the right theory of welfare, if not the theory in its entirety. So if you imagine the realm of ideas as an incredibly large mansion that we are exploring in order to find the ideas of personal welfare, we now know that at least part of the idea we are looking for is in the room labeled "desire," which we are about to enter. This is the good news from the tenuous steps taken in the previous chapter. Now for the bad news. The room we are entering is the largest room that has ever been made. It is so large, in fact, that you must pull out your telescope to verify that it is a room at all. After a great deal of effort, you are able to see the ceiling, floor, and three of the walls. However, there is no fourth wall. This, needless to say, may complicate our search, as we have a lot of conceptual ground to cover. More on this later. Let us start this chapter with a basic introduction to the fundamental desire-related concepts.

I. WHAT ARE DESIRES?

If desires, or some subset thereof, are supposed to be fundamental in determining how well your life goes for you, a solid understanding of what a desire is is in order. A desire is a pro-attitude mental state.[1] To desire

[1] I intend to use "desire," "want," and "preference" interchangeably. Although they may not be perfect synonyms, no clarity or specificity should be lost.

something is to be in favor of it, have a positive attitude toward it, give it a mental thumbs up, give it a mental "hip hip hooray!", etc.

The object of the pro attitude, or the "it" in the preceding descriptions, is a state of affairs. Accordingly, if you have a pro attitude about the state of affairs in which you dance a jig, then you desire to dance a jig.

Desires can be more or less intense. For example, your desire to dance a jig may be less intense than your desire to win the Nobel Prize. If this is true, then it should be possible to rank a person's desires, from the state of affairs that the person desires most all the way down to the state of affairs that the person most desires *not* to happen. Moreover, assigning numerical values to the ranked desires seems unproblematic, at least in theory, as long as there is nothing mysterious about claiming, for example, that one desire is twice as intense as another. The specific numbers will not matter as long as the numbers adequately represent the relative differences in intensity. In other words, if I desire the Nobel Prize ten times more than I desire to dance a jig, then, for our purposes, it will not matter if we say that the desires have intensities of 10 and 1, respectively, or intensities of 50 and 5.

II. Desire-Satisfaction Theory

Now that we have an understanding of what a desire is, we should use this information to advance the topic at hand—personal welfare. A desire-satisfaction theory of personal welfare claims, roughly, that a person's life goes better for her when the state of affairs that she desires obtains (i.e., when she gets what she wants). Along the same lines, desire satisfactionism claims that

a person's life goes worse when she does not get what she wants. The term *desire satisfaction* is used to describe the state of affairs that obtains when a person gets what she wants, and the term *desire frustration* is used to describe the result when a person does not get what she wants.

A couple points should be kept in mind regarding desire satisfactions and frustrations and their use going forward. First, I will usually speak of desire satisfactions as they pertain to making a life go better, as opposed to desire frustrations as they pertain to making a life go worse. This is, I think, a helpful consistency in framing the various issues and is also in keeping with the existing literature on this topic. However, everything that is written about desire satisfactions making a life go better is meant to apply equally to desire frustrations making a life go worse. Second, one should pay attention to the somewhat confusing nature of the terms desire satisfaction and desire frustration. While the definitions are clear, the popular use of these terms to describe *feelings* that often accompany desire satisfactions and frustrations can be misleading. So while one can *feel* satisfied or *feel* frustrated, there is no necessary connection between desire satisfactions and desire frustrations as these terms are being used here and the *feelings* of satisfaction or frustration when one does or does not get what one wants. This distinction can be seen most clearly if one takes note of the fact that desire satisfactions and desire frustrations do not require that the desirer *know* that the desired state of

affairs does or does not obtain, much less that the desires need to have any introspectively discernible feeling.[2]

Before setting out and examining a basic version of desire satisfactionism, we should be clear about another feature of these theories. Recall that objective-list theories claim that certain things are good for a person and that it does not matter what mental state, if any, the person has toward these items. Now of course, many, or most, people may want most, or all, of the items on the list. If so, the objective-list theorist will likely describe this situation by saying that the people who desire the things on the list desire them *because they are good*. Desire-satisfaction theory, on the other hand, makes a very different claim about what is good for a person. Objects are good for a person, according to this theory, *if and because* they are desired.[3] In other words, objective-list theories appeal directly to facts about value, whereas

---

[2] Some desire-satisfaction theories do have a knowledge requirement (e.g., Heathwood: 25), but this is not the majority view, nor is it a requirement for a desire-satisfaction theory. (All of the page-number-only cites in the rest of this project will be to Heathwood's forthcoming *Subjective Desire Satisfactionism* listed in the bibliography as an unpublished manuscript.)

[3] There is a growing debate in the literature about what the fundamental value bearer is in desire satisfactionism. One option is to claim that the desired object is the good thing. Another option is to claim that the desire satisfaction itself (i.e., the state of affairs consisting of both the desire for X and X) is the good thing. Although I will often write as if I am endorsing the former position, I intend to stay neutral on this issue because it appears that nothing crucial hinges on this distinction. I think both of these options could be incorporated into a fully developed theory to yield the same welfare score for all possible lives. To the extent that this is not the case, it seems as though an otherwise adequate theory could easily be modified to incorporate whichever option yields the most plausible results over a wide range of cases. For an interesting discussion of these two options, see The Two Desire-Satisfaction Views at http://peasoup.typepad.com/peasoup/2010/01/the-two-desiresatisfaction-views.html.

desire-satisfaction theories appeal only to purely descriptive facts about what people actually do (or would) want.[4] Accordingly, an objective-list theory and a desire-satisfaction theory may give the exact same personal welfare score for a particular state of affairs (e.g., a person viewing Munch's *Scream*), but the reasons for the scores will be very different.

This feature of desire satisfactionism will strike some as being completely backwards. There are people who report "experiencing value," which means that they report perceiving something as good and then wanting it for this reason. If desire satisfactionism is true, then people who experience value are simply mistaken.[5] Reports such as these do not refute desire satisfactionism, as the theory is consistent with people *believing* they experience value. It is easy to see why people would want the objects that they want to be good independent of their wanting them. This would place their desires on a solid foundation and would seem to bring all questions and doubts about what they want to an end.[6] While wanting and believing the world to be a certain way do

---

[4] See Parfit (1984: 499) for a discussion of these ideas. Also, it may be possible (although it is not clear how) for an objective-list theory to avoid appealing directly to facts about value.

[5] Here I am assuming that there is no such thing as *impersonal* value (i.e., value for the world). The existence of impersonal value would mean that goodness would exist as a mind-independent property of objects. The existence of such value would require magical metaphysics, and the experience of it would require tortured epistemology—two things I intend to avoid in this project.

[6] Adams (1999: 98) makes a similar point: "Could we sustain our valuing and enjoying if we regarded the valuing as purely subjective, merely a matter of our individual likes and dislikes? Perhaps, but it may be doubted. I suspect the

not make it any more likely that this is the case, it does raise an issue: If desire satisfactionism is true, then our desires appear to be arbitrary. I will have more to say about this issue in Chapter Three, but the short answer to this concern is that it does not matter, and even if it did, it is not true that our desires must be arbitrary. For example, it does not seem to matter that I prefer chocolate to vanilla. Make this preference as arbitrary as you like, it is simply a brute fact about me, and the fact that this is an arbitrary preference appears to have no impact on my personal welfare at all. Moreover, this is just a feature of beings who, like us, do not create themselves. I was created in such a way so as to prefer chocolate to vanilla; thus, my preference is not arbitrary after all. Finally, it is much easier to provide a plausible error theory for people who report experiencing the goodness of some objects when there is no such property than for people who do not want the good objects when there is such a property. So suppose goodness is a property of objects, and Stifler is presented with an object that has this property and is given all the information about it. However, Stifler, who wants his life to go as well as possible, does not want the object in question. What we will be forced to conclude about Stifler

---

interest in such activities as art or sport would be hard to sustain if we thought (or better, if we really felt) there was nothing more to the value of the activities and the ends we pursue in them than our liking them. It would also be hard to find meaning and interest in our own lives, more broadly if we thought that about all our activities and ends." These considerations, at most, establish that it is better for us, in terms of personal welfare, to *believe* that some of the objects we desire have objective value (i.e., value independent of our desire for them). Nothing Adams says here makes it even slightly more likely that this is actually the case. Moreover, the metaphysical and epistemological implications of Adams's objective-value world are highly suspect.

is, I think, that he is defective in some way.[7] And the greater the number of things that are supposed to have the property of goodness, the greater the number of "defective" people will become. So on the one hand, we have people incorrectly believing they experience value (which is consistent with desire satisfactionism and independently motivated), and on the other we have a potentially large number of defective people. The more plausible, and less paternalistic, conclusion is that things are good if and because they are desired and not desired because they are good.[8]

In order to properly appreciate the range and scope of the problems facing desire satisfactionism, a topic we will turn to next, it will be useful to set out a very basic version of the theory. Keeping in mind the definitions of desire satisfaction and desire frustration set out above, Unconstrained Desire-Satisfaction Theory (UDS) contains the following theses:

(i) Every desire satisfaction is intrinsically good for its subject; every desire frustration is intrinsically bad for its subject.

---

[7] If goodness is supposed to be a property of some of the objects around us, then it looks as though we will need a sixth sense in order to experience that property. If that is true, then people like Stifler could either be missing the sixth sense entirely or just have a defective sixth sense. Being forced to claim either of these does not appear to be an attractive feature of a theory.

[8] This conclusion is also more metaphysically plausible. If something like the Big Bang is the right cosmological account, then the origin of "goodness" as a property of objects looks quite difficult to explain. Of course, for a theory of personal welfare to be a theory of personal welfare, it must identify something as being intrinsically good for a subject. And this I will do, but the goodness is *for a subject*—not mysteriously slathered on objects out in the world—and the goodness will ultimately be reducible to certain facts about desires. For a good discussion of a naturalistic reduction of the normative, see Heathwood (2011b: 84-86).

(ii) The intrinsic value for its subject of a desire satisfaction = the intensity of the desire satisfied; the intrinsic value for a subject of a desire frustration = –(the intensity of the desire frustrated).

(iii) The intrinsic value of a life (or segment of a life) for the one who lives it = the sum of the intrinsic values of all the desire satisfactions and desire frustrations contained therein.[9]

UDS is just a formal way of expressing the idea that your life goes better for you when you get what you want and goes worse for you when you do not get what you want. UDS will serve as the starting point for most, if not all, of the versions of desire-satisfaction theory we will be evaluating in the pages that follow. One final note about UDS: No one has, to my knowledge, ever defended UDS as being the correct theory of personal welfare for reasons that will soon become clear.

III. WELFARE & DESIRES

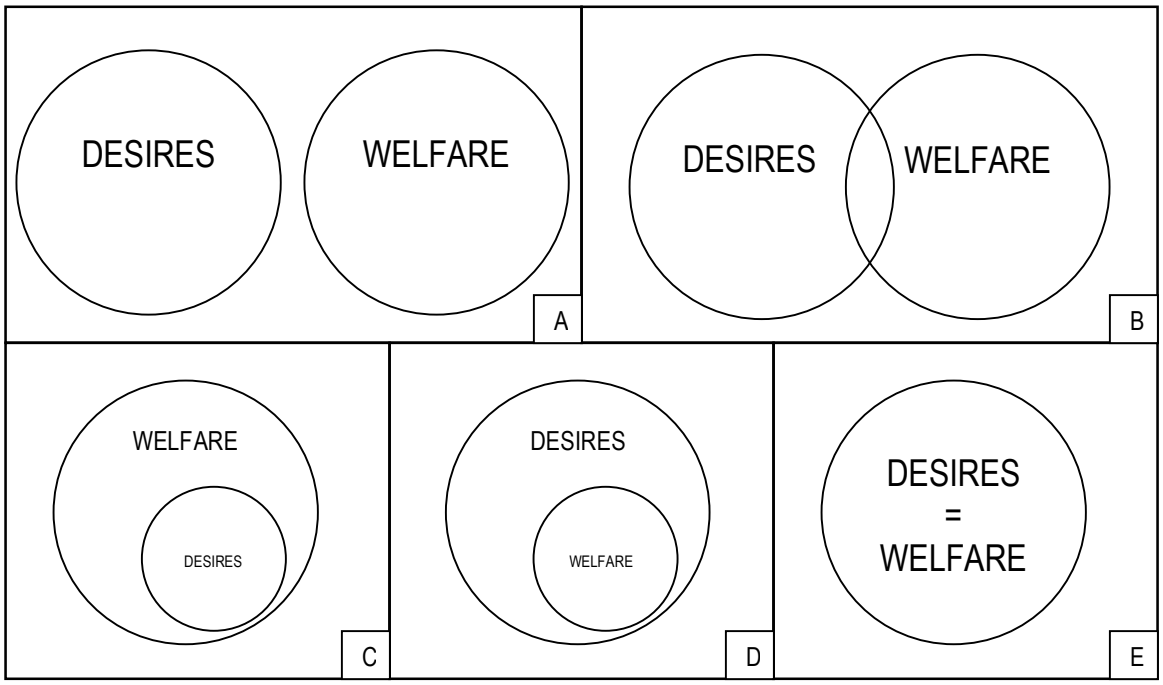
Now that we are clear about what desire satisfactionism and its related concepts are, we are in a better position to examine the relationship, if any, between welfare and desires. This examination will illuminate the general types of objections that are leveled against the theory as well as the logical space that a desire-satisfaction theorist must defend. The graphic below, designed to aid in this examination, depicts the five logically possible relationships between welfare and desire.[10]

---

[9] This is a modified version of a desire-satisfaction theory that appears in Heathwood (2005: 489).

[10] The graphic does not contain Venn diagrams in the technical sense. For example, Diagram E is only meant to depict an intimate connection between desires and welfare rather than an identity relationship.

Diagram A depicts a scenario in which desires and welfare are completely distinct concepts, and neither impacts, or is impacted by, the other. For those interested in establishing the truth of Diagram A, the typical strategy is to provide an alternative theory of welfare rather than to formulate objections to desire satisfactionism. After all, formulating an objection to desire satisfactionism that purports to show *all* desires are not relevant to a person's welfare is a tall order, and an objection that



establishes only a certain subset of desires does not affect personal welfare gets you only a small part of the way, at best, to establishing the truth of Diagram A. The alternative theories, as we saw in the last chapter, usually come in the form of some version of a hedonistic or objective-list theory. However, these theories, to the extent they exclude all desires, are false. As was demonstrated in the last chapter, personal welfare entails caring and caring entails desire.

Therefore, personal welfare entails desire. Diagram A does not accurately portray the relationship between welfare and desires.

Diagram B depicts a scenario in which some desires affect welfare, and some things other than desires also affect welfare. This possibility merits serious consideration for a couple reasons. First, this option has a great deal of intuitive plausibility. In fact, I suspect this is the diagram most people would choose as being the correct depiction of the relationship between welfare and desire. This fact (if it is a fact), of course, does not carry a great deal of philosophical weight in and of itself, but I think ideas that might not otherwise merit a great deal of scrutiny should be looked at more closely the more they are believed by intelligent, but philosophically uninitiated, people. Second, and much more importantly for our purposes, Diagram B is not ruled out by the latest version of our Internalist Principle, IP'. Recall that IP' states that the value of a life (or part of a life) for the one who lives it is determined to a significant degree by what the person in question cares about. Accordingly, all we have established so far is that welfare is determined *to a significant degree* by desires (since caring entails desire). This obviously leaves open the possibility that there are determinants of welfare other than desires, which is exactly what Diagram B depicts.

Since the target concept is personal welfare, objections to Diagram B will come in the form of a claim that something other than desires completely determines how well a life goes for the person who lives it. Not having addressed this sort of objection in the last chapter, it is time to do so here. In

addressing this objection, I intend to cover all of the logical space in the non-desire area of the welfare circle in Diagram B. This will be accomplished in two steps, first by addressing the possibility of pleasure as a non-desire-related determinant of welfare, and then by addressing the possibility of items normally included in objective-list theories (or any other item for that matter) that might fill this role.

IV. A Desire/Pleasure Hybrid Theory?

One might plausibly claim that certain pleasures do not make a life go better for the person who lives it and that certain pains do not make a life go worse. For example, it may be argued that a refreshing sip of iced tea or a mildly stubbed toe fits this description. However, whether these claims are plausible or not, there is a stronger claim about pleasure and pain that lacks any plausibility at all. This is the claim that all pleasures and all pains do not, and in principle cannot, affect personal welfare to any degree. That pleasure and pain have an effect on personal welfare is a data point that a welfare theorist ignores at his professional peril. To claim that pleasure and pain are irrelevant to prudential value is to embark on what seems to be one of the longest uphill climbs a philosopher could undertake.

The question, then, as it relates to Diagram B, is where does this data point belong? It clearly belongs in the welfare circle, but does it belong in the part that overlaps with the desire circle or in the non-desire part of the welfare circle? The somewhat surprising answer is that the pleasure data point belongs in the desire circle. This view, called the *Motivational Theory of*

Pleasure (MTP), might require an entire book to thoroughly defend.[11] Consequently, all I will have space to do here is to present and motivate the view. Needless to say, if MTP turns out to be false, then that fact would prove to be a fairly serious issue for the theory that I ultimately present and defend in Chapter Five (just as it would be for any other desire-satisfaction theory).

In order to appreciate the attractiveness of MTP, a cursory examination of its main competitor, *Felt-Quality Theory* (FQT), should prove useful. Carson (2000: 13) describes FQT as holding "that the pleasantness or unpleasantness of an experience is determined solely by its felt introspectable or phenomenological qualities." FQT is easy to understand, probably has the most initial plausibility of any theory of pleasure and pain, and has been endorsed by a number of philosophers.[12] Two problems for FQT will help to highlight the advantage of MTP as an alternative. First, there is the *heterogeneity problem*, which is nicely illustrated by the following:

> There are bodily pleasures, like those had from relaxing in a Jacuzzi tub, from sunbathing on a warm beach, or from sexual activities. There are gustatory and olfactory pleasures (maybe they, too, qualify as "bodily"). There are what we might call "emotional pleasures," such as the elation of receiving an ovation or the prideful satisfaction of completing a difficult and worthwhile project. There are more "cognitive" pleasures, such as the pleasure derived from working on a crossword puzzle, from reading an insightful philosophy paper, or from listening to an amusing anecdote. There are aesthetic pleasures, like those derived from listening to beautiful music or from taking in a powerful sculpture. (Heathwood 2007: 25)

---

[11] Heathwood (2007) devotes a paper to defending MTP only as it relates to sensory pleasure.

[12] See, e.g., G. E. Moore (1903: § 12), and C. D. Broad (1930: 225-31).

If the proponent of FQT admits that these all count as pleasures, which it looks like he must, then he has a problem. According to FQT, something is a pleasure in virtue of a felt quality. If all the things on this list are pleasures and FQT is the right theory of pleasure, then there must be some felt quality that each of these pleasures shares, in virtue of which they are actually all pleasures. However, a little reflection on the phenomenology of these experiences should reveal that there is no felt quality that they all share.[13] Therefore, FQT is false.

The *oppositeness problem* is the second way in which FQT fails. Pleasure and pain are often described as being opposites.[14] If this is true, then FQT should be able to explain this fact. It is unclear whether FQT cannot, in principle, explain this fact or whether it merely does not do so. Getting to the bottom of this issue may require a long foray into philosophy of mind that would be an unnecessary digression here. After all, FQT needs to be false only once. At any rate, it is conceptually difficult, if not impossible, to see how one felt quality could be the opposite of another felt quality, no matter how much one tried to finesse this result by choosing the easiest possible case.

The Motivational Theory of Pleasure, on the other hand, does not suffer from these problems. It nicely accounts for the common element in pleasures

---

[13] This point has been made by several philosophers. See, e.g., Sidgwick (1907: 127), Feldman (1997: 87), and Sobel (2002: 241).

[14] This is also of great concern to hedonistic theorists in order to prevent commensurability problems whereby if pleasure and pain are not opposites, they cannot be weighed against each other on the same scale.

and pains and how pleasure can be the opposite of pain. After noting the heterogeneity problem, Parfit (1984: 493) writes:

> What pains and pleasures have in common are their relation to our desires. On the use of 'pain' which has rational and moral significance, all pains are when experienced unwanted, and a pain is worse or greater the more it is unwanted. Similarly, all pleasures are when experienced wanted, and they are better or greater the more they are wanted.

In addition to Parfit, similar versions of MTP have been endorsed by Alston (1967: 345), Brandt (1979: 38), Carson (2000: 13), and Heathwood (2007: 24). While the differences in the various formulations of MTP are interesting and worthy of examination in their own right, they need not detain us here. The only thing required for our purposes is that a version of MTP is the correct theory of pleasure such that pleasure is reducible to the more basic concept of desire (i.e., pleasure is reducible to desire if facts about pleasure just are facts about desire). If a version of MTP is true, then, as noted by Parfit above, we have an answer to the heterogeneity problem: The common element in all pleasures, and what in fact makes something a pleasure, is that it is desired.[15] Also, the oppositeness problem now has an answer: Pleasure is desired, pain is desired not, and desire and desire not are opposites (Heathwood 2007: 27). Finally, and most importantly, the truth of MTP means that the pleasure/pain data point belongs in the area on Diagram B where the desire and welfare circles overlap.

---

[15] For all claims about pleasure as it relates to MTP, a corresponding claim could be made about pain and is meant to be implicit.

V. A DESIRE/OBJECTIVE-LIST HYBRID THEORY?

There remains the possibility, as depicted by Diagram B, that something that is not desire, nor reducible to desire (as we just showed pleasure to be), can affect welfare. I will call theories of this type *Desire/Objective-List Hybrid Theories*, although in this category I will evaluate the prospect of any non-desire-based object affecting welfare regardless of whether it has, or ever will be, included in an objective list.[16] In order to properly evaluate these theories, an examination of various formulations will be beneficial. Parfit (1984: 501-02) ends his discussion of personal welfare in *Reasons and Persons* by saying that, although he will not attempt to answer the question of which theory of prudential value is correct, he will "mention" another theory, "which might be claimed to combine what is most plausible in these conflicting theories."

> We might claim, for example, that what is good or bad for someone is to have knowledge, to be engaged in rational activity, to experience mutual love, and to be aware of beauty, while strongly wanting just these things. On this view, each side in this disagreement saw only half of the truth. Each put forward as sufficient something that was only necessary. Pleasure with many other kinds of object has no value. And, if they are entirely devoid of pleasure, there is no value in knowledge, rational activity, love, or the awareness of beauty. What is of value, or is good for someone, is to have both; to be engaged in these activities, and to be strongly wanting to be so engaged.

---

[16] A Desire/Objective-List *Pluralist* Theory (two kinds of things are good for us—desire satisfactions and things on an objective list) would not make Diagram B true, as this would essentially be just a form of an objective-list theory since these theories claim that at least *some* of the things that are good for us are objective. Moreover, a theory of this type would be plagued with all of the problems associated with objective-list theories and all of the problems associated with desire-satisfaction theories—a tough row to hoe.

Adams (1999: 93-94), in *Finite and Infinite Goods*, writes:

> Without pretending to offer here a complete theory of the nature of a person's good, I wish to explore the idea that what is good for a person is a *life* characterized by *enjoyment of the excellent*. More precisely, I shall argue that the principal thing that can be noninstrumentally good for a person is a life that is hers, and that two criteria (perhaps not the only criteria) for a life being a good one for a person are that she should enjoy it, and that what she enjoys should be, in some objective sense, excellent. Its being more excellent, and her enjoying it more, will both be reasons for thinking it better for her, other things being equal . . . .

Darwall (2002: 80) writes in *Welfare and Rational Care*:

> The specific version of the Aristotelian Thesis I shall defend, then, is that the most beneficial human life consists of activities involving the appreciation of worth and merit. I do not claim that appreciating these values is the only source of human good. I only claim, somewhat vaguely, that is the most important source.[17]

Finally, Feldman (2004: 120), in *Pleasure and the Good Life*, formulates "Desert-Adjusted Intrinsic Attitudinal Hedonism" (DAIAH), which he describes as follows:

> The idea is to say that the intrinsic value of an attitudinal pleasure is determined not simply by the intensity and duration of that pleasure, but by these in combination with the extent to which the object of that pleasure deserves to have pleasure taken in it. More exactly, the value of a pleasure is enhanced when it is pleasure taken in a pleasure-worthy object, such as something good, or beautiful. The value of a pleasure is mitigated when it is pleasure taken in a pleasure-unworthy object, such as something evil, or ugly. The disvalue of pain is mitigated (the pain is made less bad) when it is pain taken in an object worthy of pain, such as

[17] To clarify his claim both in general and with respect to ensuring that it is about personal welfare, Darwall (2002: 76) writes: "My claim will be that a person's welfare is enhanced, her life is made better *for her*, through active engagement with and appreciation of values whose worth transcends their capacity to benefit (extrinsically or intrinsically). The benefit or contribution to welfare comes through the *appreciative rapport* with the values and the things that have them."

> something evil, or ugly. The value of a pain is enhanced (the pain is made yet worse) when it is pain taken in an object unworthy of this attitude, such as something good or beautiful.

The first thing to notice about these theories is that they will all be subject to the same objections that were briefly mentioned in the last chapter as they applied to straight objective-list theories.[18] This is because, as is the case with objective-list theories, some things are claimed to have "value[] whose worth transcends their capacity to benefit" (Darwall 2002: 76). This means that the authority objection still applies: Who decides what goes on the list and what, if anything, do these items have in common with one another? Also, the commensurability objection is still a problem: How does the value of each item on the list compare to that of the other things on the list? Is this even possible according to the theory? Finally, the completeness objection still must be answered: Is this list of items complete?

The other noteworthy feature of these four theories is their inchoate nature, which is the reason that an extended quote from each of them appears above.[19] Parfit (1984: 501-02) "mentions" what such a theory "might claim,"

---

[18] Or, perhaps, the first thing to notice about these four theories is that they all seem to be pleasure/objective-list hybrid theories rather than instances of desire/objective-list hybrid theories that I claimed to be addressing. Two points should be made here. First, as I just argued, pleasure is reducible to desire. Therefore, each of these theories can easily be interpreted as desire/objective-list theories. Second, recall that the main objection leveled against objective-list theories dealt with the problem of people not caring about the items on the list. Interpreting these theories as desire/objective-list hybrid theories is the most charitable reading that will perhaps allow them to survive this objection, which will be evaluated below.

[19] Kraut (1994: 44) offers yet another incomplete theory of this type:

but does not flesh out the theory in the single paragraph he devotes to it. Adams (1999: 93), in setting out his theory, does not even "pretend[] to offer here a complete theory of the nature of a person's good." Darwall (2002: 80) also admits to offering an incomplete theory and claims "somewhat vaguely, that [appreciation of worth and merit] is the most important source [of human good]." Finally, Feldman (2004: 121) offers a complete theory only in the sense that it purports to include everything that determines prudential value.[20] While he does briefly defend the use of the concept *desert* in modifying his base hedonistic theory to meet a specific objection, he does not flesh the concept out.[21] It is barely worth mentioning that it is hard to properly evaluate a theory with an unexplained central concept.

> So, there are at least three conditions that make a life a good one: one must love something, what one loves must be worth loving, and one must be related in the right way to what one loves. Perhaps other conditions must be specified, but I will not explore that possibility here.
>
> It might be objected that the thesis I am proposing is empty unless it is backed by a systematic theory that enables us to decide which among alternative ways of life is most worth living and which objects are most worth loving. It would of course be nice to have such a theory, but it is possible to do without one and still make defensible judgments about what is worth wanting and what is not.

[20] Feldman's (2004: 121) exact quote: "The intrinsic value of a life is entirely determined by the intrinsic values of the episodes of intrinsic attitudinal pleasure and pain contained in that life, in such a way that one life is intrinsically better than another if and only if the net desert-adjusted amount of intrinsic attitudinal pleasure in the one is greater than the net amount of that sort of pleasure in the other."

[21] Feldman (2004: 122) notes that "it would be good to have a fully developed theory of desert," but "that's a project for another book."

What, then, can we safely surmise here? One possibility is that a desire/objective-list hybrid theory is much more plausible than it seems as presented here because I have simply cherry-picked underdeveloped theories from minor philosophers. This is not the case. While this list is probably not exhaustive, I have searched for a complete desire/objective-list hybrid theory to no avail. Also, each of these philosophers is quite well known in philosophy circles generally and in axiology specifically. In fact, a book on personal welfare that did not mention each of these theorists would be suspect. The other main possibility, then, is that this type of theory is just not very plausible. These theories sound noble and seem plausible at a sufficiently abstract level, which probably explains why none of them appear to get beyond this stage. However, this type of theory is not plausible as a theory about what makes a life go best *for the person living it*. This point always bears repeating as, I think, many theorists lose sight of the target along the way and end up giving, unwittingly, a theory of something else entirely.[22] A concrete example, of the sort I did not find in any of the four sample theories, should help illuminate this weakness in theories of this type.

I love movies and have watched thousands of them. It is a very widely held belief that *Citizen Kane* is the greatest film ever made. In fact, this might be the most widely held belief there is regarding the greatest example of a work of art in any artistic genre. I do not enjoy watching *Citizen Kane* very much, although I will watch it as an academic exercise. On the other hand, *Night of*

[22] I tend to think of desire/objective-list hybrid theories as a theory of what sort of friends the theorist would most like to have.

the Comet, a 1984 valley girl/zombie B-movie, is a movie that I watch often and thoroughly enjoy every time. Assuming that each desire/objective-list hybrid theory will claim that *Citizen Kane* is (much?) more worthy of pleasure/desire than *Night of the Comet*, then some variations of this sort of theory will claim that my life goes better for me by watching *Citizen Kane*. This is the wrong result because it seems to clearly violate PCP (i.e., there does not appear to be any theoretical interest, compelling or otherwise, to make a paternalistic claim here) and it threatens to violate IP' (i.e., if this sort of claim is allowed to stand, then there seems to be nothing stopping similar claims, which could easily lead to the value of a life being determined to a significant degree by things that the person cares *nothing* about).

This example obviously does not entirely refute this type of theory. We could weaken the objective element to such an extent as to provide the right result in my movie case. While I suspect that most of the desire/objective-list hybrid theorists will think that weakening the objective element to achieve the right result in a wide range of these sorts of cases will weaken the theory too much for it to achieve their desired results in other cases, this is a viable strategy. However, problems remain. Consider a variation of the movie case presented above. Spike enjoys *Citizen Kane* just as much as I enjoy *Night of the Comet*. If we sit across the table from each other, he watching his movie and I watching mine, what would be the basis for claiming that Spike's life is going better for him than my life is going for me for that stretch of time? I understand the *argument* in favor of this conclusion, but here I mean to call

into question the *basis*, or, more accurately, the metaphysical foundations, of the theory.

In order to be a complete theory of personal welfare, a theory must be able, in principle, to give a definitive answer to how well a life went for the person who lived it, thus making a ranking of possible and actual lives possible. For desire/objective-list hybrid theories, this means that the theory will have to account for how every possible object of desire comes to have its value, or lack thereof. There are several issues confronting the metaphysics of these theories. First, the property of "good for a person to get" is just an odd sort of property for an object out in the world to possess. A second related issue deals with the timing of the acquisition of value; at what point does the object acquire the value that it has? Third, these theories must work out the relative weight they give to desires versus the weight they give to the objects. For example, can a relatively strong desire for a nearly worthless object outweigh a weak desire for an excellent object? Fourth, the theory must be able to work out the relative weights *between* objects. So is *Citizen Kane* twice as worthy of desire as *Night of the Comet*? Ten times? One hundred? And how does *Citizen Kane* stack up against Picasso's *Guernica*, Beethoven's *Ninth Symphony*, and Michelangelo's *David*? Although there are also metaphysical issues, it would be beating a dead horse to pursue them here. And even if the metaphysical issues were thoroughly explained, there would still be all of the related epistemological issues. These all count as reasons, in addition to the

general implausibility of these theories, why, I think, none of these theories are complete.[23]

The last objection to desire/objective-list hybrid theories involves a final range of implausible consequences from different scenarios in which a person gets everything he desires. Other things being equal, one would probably suppose that such a person had at least a good life and perhaps one of the best possible lives. However, desire/objective-list hybrid theories will probably have to claim (remember, none of them are complete) that some lives in which the person gets everything he desires are either equivalent to having no life at all or, in some cases, worse than having no life at all. This consequence stems from the fact that these theories have, as an essential element, ideas of objects that are "excellent," "worthy of being enjoyed," or "pleasure-worthy." Accordingly, for such a theory to be plausible, it must claim that some objects possess none of this property. If so, then a life filled with these objects and the desire for them is equivalent to, in terms of intrinsic value, never having lived at all. Perhaps my desire for *Night of the Comet* is like this. Or perhaps it is even worse than this, such that the enjoyment of this movie makes my life go worse for me. This would be the case if *Night of the Comet* has a negative "excellence" value. Even if it does not, it seems quite plausible that a potentially great number of things could have a negative value. Taking into

[23] I think one of the reasons there does not appear to be even one complete version of this type of theory in the over two-millennia history of philosophy is that a complete theory (i.e., one that had answers to all of the questions I have raised) would make the general implausibility of this approach self-evident. In other words, these theories retain any semblance of plausibility precisely because they are incomplete.

account the vast range of objects in the world, it would seem odd, after all, if the continuum of excellence values was limited on the low end by zero.[24] If the values can go negative, then according to these theories a life filled with desire satisfactions could be *worse* than no life at all. This is not plausible. Desire/objective-list hybrid theories should be rejected for all of these reasons.[25]

VI. Taking Stock & Challenges on the Road Ahead

By evaluating objective-list and hedonistic theories of welfare, we determined that a person's welfare is determined to a significant degree by what the person cares about (as stated in IP'). After determining that welfare entails caring and that caring entails desire, we evaluated the possibility of something in addition to desires being a determinant of welfare. This turned out not to be the case. Accordingly, a person's desires (or some subset thereof) are the sole determinants of a person's welfare.

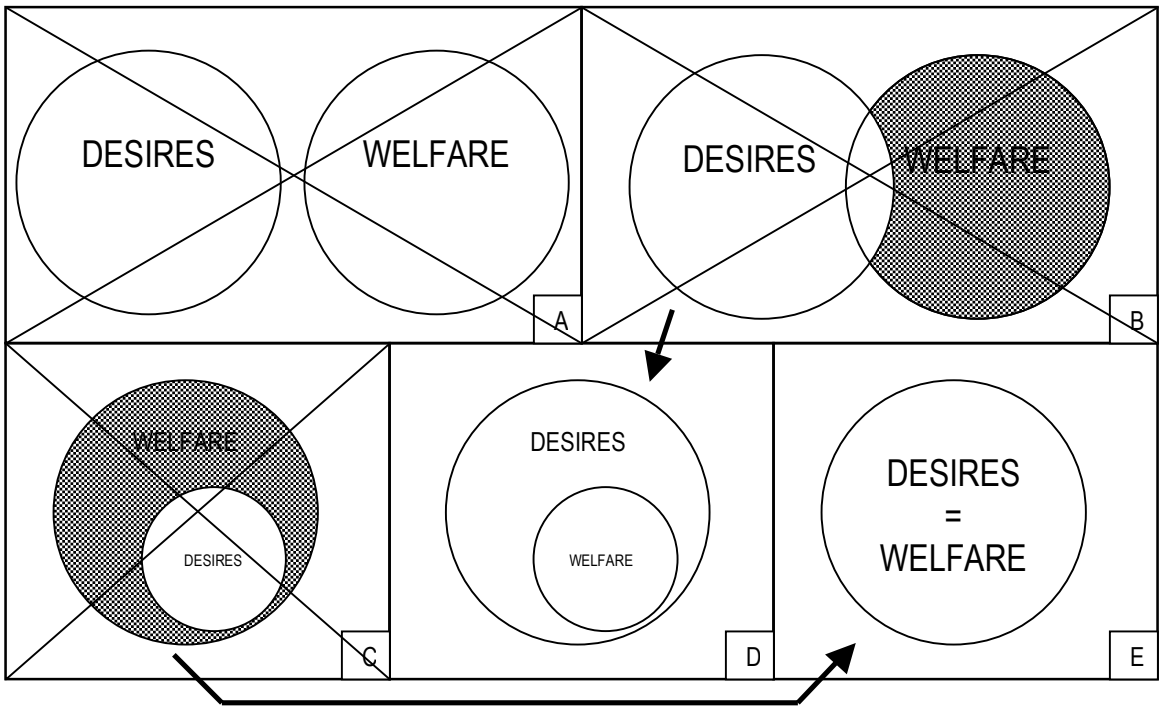
In terms of where this leaves us on our graphic depicting the logical possibilities, we can now cross off Diagrams B and C. Stated another way, by

---

[24] Such a claim could only be made by a theorist who had never visited Rotten.com or been subjected to *Two Girls One Cup*.

[25] There is, of course, one final reason to reject these theories—the fact that there do not appear to be any complete versions of such a theory. In this era of prolific academic philosophy, this fact alone should cast doubt upon the plausibility of these theories. It is fairly easy to make any number of false theories look good at a sufficiently high level, but the devil is always in the details. A defender of such views may claim that all of the above objections do not apply to *her* preferred version of the theory. But no such theory exists today. And I think the future prospects for a version of this theory that is both plausible and that avoids these objections are very, very dim.

eliminating the possibility of a non-desire welfare determinant, we can shade out those areas in Diagrams B and C. After shading out these areas in each diagram, both of them become slightly different graphic depictions of our two remaining options. Shading out the non-desire area of the welfare circle in Diagram B is just another way of depicting Diagram D, which is one of our remaining possibilities. Similarly, shading out the non-desire area of the welfare circle in Diagram C is just another way of depicting the other remaining option, Diagram E.



Narrowing the possibilities for a theory of personal welfare to the two depicted above is an achievement in itself and should give us reason to think that we are getting close to the correct theory. Knowing where to look, after all, usually makes the finding easier. However, this is not as great of an

accomplishment as it might appear at first blush for a couple of reasons. First, desire-satisfaction theories of welfare are the dominant view, and their pedigree stretches a long way back into the history of ideas.[26] We stand at the same spot where many have stood before. Second, for reasons that will soon become abundantly clear, the most difficult stretch of the journey starts here. To understand why this is the case, a preview of the road ahead will be helpful. Not surprisingly, the diagram that depicts the correct relationship between desires and welfare is Diagram D. What makes this the case is that we will easily find at least one type of desire that will not affect welfare, and as it turns out, we will find several types of desires that do not affect welfare. The real accomplishment, then, will be delineating the ambit of desires that do affect welfare. Depicted graphically, this will be the project of shading out the non-

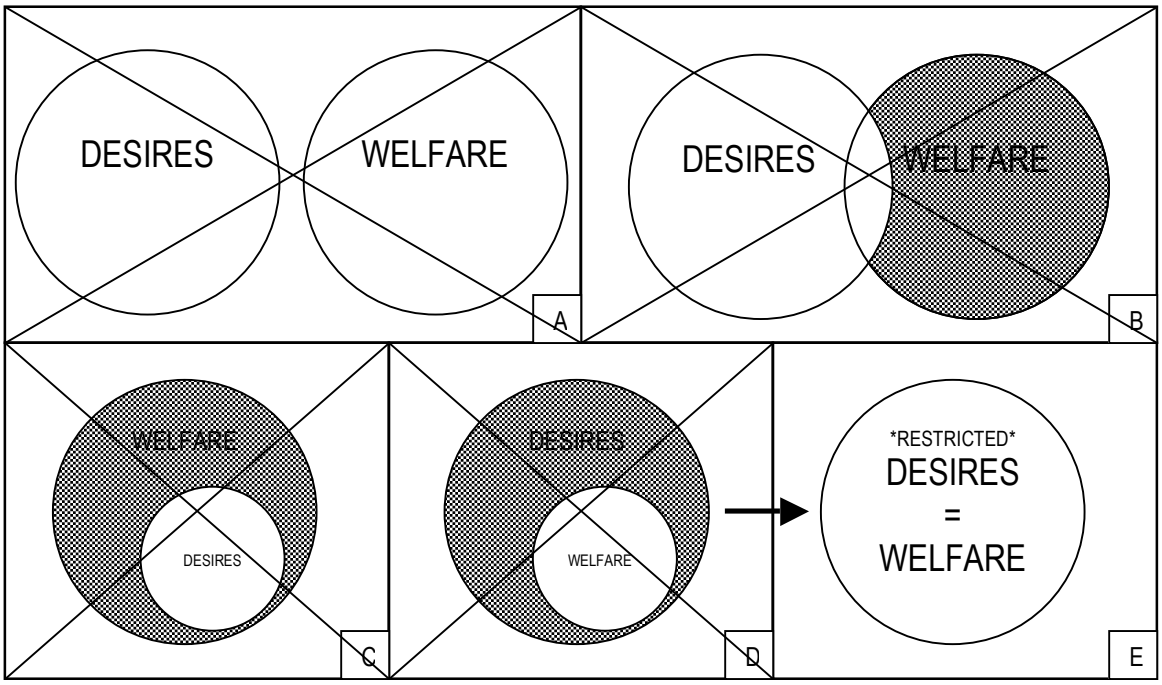
---

[26] Hobbes (2002: 42), in 1651, makes a remark that is at least consistent with desire satisfactionism, if not an outright endorsement:

> But whatsoever is the object of any man's appetite or desire, that is it which he for his part calleth *good*; and the object of his hate and aversion, *evil*; and of his contempt, *vile* and *inconsiderable*. For these words of "*good*, *evil*, and *contemptible* are ever used with relation to the person that useth them, there being nothing simply and absolutely so, nor any common rule of good and evil to be taken from the nature of the objects themselves . . . .

Spinoza, in 1677, makes several remarks that could be characterized in the same way. For example: "It is thus plain from what has been said, that in no case do we strive for, wish for, long for, or desire anything, because we deem it to be good, but on the other hand we deem a thing to be good, because we strive for it, wish for it, long for it, or desire it" (Spinoza 1981: 118), and: "For I have shown that we in no case desire a thing because we deem it good, but, contrariwise, we deem a thing good because we desire it . . ." (Spinoza 1981: 136). The earliest modern discussion of desire satisfactionism in academic philosophy came in 1874 from Sidgwick (1907: 111-12): "[A] man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time."

welfare area of the desire circle in Diagram D. Successfully accomplishing this goal will mean that we have accurately defined the subset of classes that affect welfare, which will move us from Diagram D to a slightly modified version of Diagram E that now reads "Restricted Desires" to reflect the fact that we are now dealing only with a subset of desires after the move from the successfully shaded Diagram D.



Once we have accomplished that task, we will be in a position to fill in the details of the theory in order to show how, exactly, our restricted desires affect personal welfare. Finally, the advantages of our new desire-satisfaction theory over competing theories will be demonstrated, as well as how the theory handles a variety of objections made to desire-satisfaction theories. Before

tackling these topics, however, the rest of this chapter will be devoted to chronicling the problems that desire-satisfaction theories face.

VII. Some General Concerns About Desires as the Basis for a Theory of Welfare

Suppose, again, you are told that a person had gotten everything she desired during her life, and then you are asked to venture a guess as to how well her life went for her. If you are not told anything more about her life, you would probably guess that her life went somewhere between well and very well for her. This seems to be in accordance with how we generally view desire satisfactions; when we learn that someone got what she wanted, we assume that, other things being equal, her life is now going better than it was before. In other words, we are using desire satisfactions as a *proxy* for personal welfare.

Desire-satisfaction theories of welfare, however, make a much stronger claim. These theories do not claim that certain desire satisfactions are simply a proxy for welfare. Rather, they claim that the satisfaction of the desires included in the theory is what contributes to welfare and that nothing other than these desires affect welfare at all. This strong of a claim is almost always accompanied by a host of objections that it must overcome before it will be generally accepted. The first set of three objections—considered in the next few paragraphs—applies to all classes of desires.

While the next two objections deal with the *objects* of our desires, the first one, fittingly, deals with the *origin* of our desires. Nietzsche claimed that

philosophers do not like to talk about the beginnings of things.[27] I suspect there are many reasons for this, but I think the main reason pertains to the long shadow cast by the single most impenetrable, inscrutable question ever asked: Why is there something rather than nothing?[28] Anyone who pretends to offer an answer to this question is a fool of unparalleled stature.[29] So if the origins of things is a generally underexplored and difficult topic, it should come as no surprise that the origins of our desires have been neglected as well. Given what follows, it is clear why proponents of desire satisfactionism should want to avoid this topic, but it is not clear at all why opponents of such

---

[27] Nietzsche makes this point, like most of the points he makes, repeatedly. For example: "You want to know what the philosophers' idiosyncrasies are? . . . Their lack of historical sense for one thing, their hatred of the very idea of becoming, their Egypticity. They think that they are showing *respect* for something when they dehistoricize it, *sub specie aeterni*, – when they turn it into a mummy" (Nietzsche 2005: 166-67). See also Nietzsche (1996: § 2).

[28] Neither god nor the Singularity of the Big Bang is an answer to this question, as both of these things are *things* that would require an explanation in order to answer the question.

[29] David Hume (1989: 32-33) summed this up best: "A very small part of this great system, during a very short time, is very imperfectly discovered to us; and do we thence pronounce decisively concerning the origin of the whole? Admirable conclusion!" Bede Rundle (2004: ix), apparently undeterred, wrote an entire book, *Why There is Something Rather than Nothing*, on this topic and came up with, among other things, the "central thesis that there has to be something." Rundle (2004: ix) admits he cannot claim that the arguments on key points are "compelling," but rather claims that "good reasons can be given in support of the position advocated." It is not at all clear what that means, but what is clear is that Rundle does not understand the essence of the question, as he appears to offer a linguistic analysis as an argument for nothingness being impossible (i.e., there *has* to be something). It is precisely this sort of thing that gives philosophy a bad name.

theories would do so.[30] This is because the origins of a great many of our desires come from a place that is both difficult in its own right to analyze and seems very unreliable as a source of things designed to make our lives go better. The place that I am referring to is, of course, the parts of our psychic makeup that trace back to our animal origins. Humans are the product of billions of years of evolution, and, for better or worse, we carry that baggage with us to this day. And while that baggage has made the human race quite good at *survival*, that is not what concerns us here. Our topic, is, roughly, what makes us *thrive* as a species, rather than what makes us merely survive.

---

[30] One critic, Adams (1999: 90), does touch upon this problem: "And [desire satisfactionism] still is not plausible if we also demand full realization of the causal history of one's motives. A fuller understanding of our histories, capacities, and tendencies does not always make us like ourselves better." Rawls (1999: 368-69) also discusses some concerns regarding the origin of desires:

> We may also investigate the circumstances under which we have acquired our desires and conclude that some of our aims are in various respects out of line. Thus a desire may spring from excessive generalization, or arise from more or less accidental associations. This is especially likely to be so in the case of aversions developed when we are younger and do not possess enough experience and maturity to make the necessary corrections. Other wants may be inordinate, having acquired their peculiar urgency as an overreaction to a prior period of severe deprivation or anxiety. The study of these processes and their disturbing influence on the normal development of our system of desires is not our concern here. They do however suggest certain critical reflections that are important devices of deliberation. Awareness of the genesis of our wants can often make it perfectly clear to us that we really do desire certain things more than others. As some aims seem less important in the face of critical scrutiny, or even lose their appeal entirely, others may assume an assured prominence that provides sufficient grounds for choice. Of course, it is conceivable that despite the unfortunate conditions under which some of our desires and aversions have developed, they may still fit into and even greatly enhance the fulfillment of rational plans. If so, they turn out to be perfectly rational after all.

A cursory glance at the human condition, both current and historical, should sharpen one's appreciation for this critical distinction. Accordingly, the fact that a great number of our desires simply bubble up, for lack of a better term, to our conscious minds seems to be a significant issue for desire-satisfaction theories to address.

The questionable origin of many of our desires contributes to the second major issue for desire satisfactionism—the fact that we can be *mistaken* about our desires. If some of our desires do just bubble up to our conscious minds, then it seems plausible to suppose that some of our desires stay down in the primordial muck. And if Freud was right, this is precisely what happens. This opens the door to several problematic possibilities. For example, we could want X and be unaware of it, we could want X and think we do not want X, we could not want X and think that we do want X, etc. I think it is both more tempting and more plausible for a desire-satisfaction theory to sidestep these issues than the origin issue, but avoiding this issue does leave the theory open to a line of objections seeking to exploit this fact.

The final general issue with desires "is that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one could take as touching one's own well-being" (Griffin 1986: 17). While that sounds daunting enough from the perspective of desire satisfactionism, Griffin actually understates the problem or, at least, does not emphasize the proper scope of the issue. This is because it is possible to desire *anything*, where "anything" has a meaning that extends

far beyond the boundaries of its meaning in any other context. This issue is what I was alluding to in the beginning of the chapter when I claimed we had found the room in the mansion that contains welfare, but that the room was not only massive but also lacking a fourth wall. Starting with the easiest to imagine and working our way up, here are some examples that illustrate the scope of the problem. First, there are no *ethical* limits on what we *can* desire. As Hume (2003: 296) says, I could "prefer the destruction of the whole world to the scratching of my finger." Second, there are no *epistemic* constraints on what we can desire. In other words, it need not be possible for us to *know* about some state of affairs in order for us to want it. I could, for example, desire that the creature in the universe that is most similar to me have a good life provided that this creature does not live on Earth or on any planet that I could possibly ever visit or communicate with. I could desire to experience the infinite. I could also desire to experience, literally, *everything* at once. Third, there are no *metaphysical* constraints on what we can desire.[31] "For example, I may want it to be true that, in my drunkenness last night, I did not disgrace myself" (Parfit 1984: 171). Suppose that I did disgrace myself last night. I now want something that is metaphysically impossible to achieve. Finally, there are no *logical* constraints on what we can desire. "The Pythagoreans wanted the square root of two to be a rational number. It is logically impossible that this desire be fulfilled" (Parfit 1984: 172).

---

[31] Nagel's (1970: 76) "man who wastes his life in the cheerful pursuit of a method of communicating with asparagus plants" is another amusing example of an impossible desire.

There are a potentially infinite number of variations of the types of desires described in the last paragraph, and this does not even include all the variations of the more mundane desires that are a part of everyone's daily lives. The issue that some may have with the majority of the infinite list of possible desires is a nagging doubt about whether someone could actually want these things. There are, of course, states of affairs that no one has ever wanted, but the point is that someone *could* want them. And he could want these states of affairs even if an omnipotent god could not make them obtain (Parfit 1984: 172). In short, it is a logical truth that we want what we want, ethics, epistemology, metaphysics, and logic notwithstanding. Although this leaves a lot of conceptual ground for a desire-satisfaction theory to cover, there is reason to be optimistic about the chances of a solution. Narrowing the list of potential desires to those that affect welfare is much less theoretically daunting than either expanding a hedonistic theory or creating an objective-list theory out of whole cloth.

VIII. Defective Desires

As mentioned above, Diagram D will be the correct depiction of the relationship between desires and welfare if we can find just one desire, or perhaps even an entire class of desires, that does not affect welfare. It is also possible that there are desires that do affect welfare, but do not fall within the welfare circle in Diagram D. In other words, there may be desires the satisfaction of which makes a life go *worse* for the person who lives it. The correct desire-satisfaction theory of welfare will have to address each of these

possibilities. The basic form of the argument made against desire theories of welfare is usually as follows:

> *Premise 1:* Desire-satisfaction theories include in the calculation of personal welfare the satisfaction of desires of type X.
> *Premise 2:* The satisfaction of desires of type X do not make a life go better.
> *Conclusion:* Therefore, desire-satisfaction theories of personal welfare are false.

The goal, then, for the rest of this project is to develop a new desire-satisfaction theory, motivate it, and defend it against this type of objection.

The rest of this section will be devoted to cataloguing the types of desires that have been advanced in the literature on personal welfare as posing a threat to the success of desire satisfactionism. In fact, the existence of one or more of the types of desires listed in this section has been thought to be a decisive objection against desire satisfactionism by Brandt (1979: 115-26), Schwartz (1982: 195-97), Griffin (1986: 10), Kraut (1994: 40), Kagan (1998: 38-39), Adams (1999: 87), Carson (2000: 80), and Feldman (2004: 16).

Parfit (1984: 494) discusses the scenario of meeting an ill stranger and desiring that the stranger be cured. Although Parfit never knows about it, the stranger is subsequently cured. Parfit asserts that it is not plausible to claim that the satisfaction of this desire makes his life go better. The rationale for this claim is that this desire, and many other possible desires, is too *remote* from Parfit in some way to impact his welfare. The possibility of remote desires is a direct consequence of the unbounded nature of our desires.

Kraut (1994: 41) asks us to imagine a boy who desires, on *impulse*, to throw a rock at a nearby duck. Kraut suggests that the satisfaction of this

desire does not increase the boy's welfare and that it may even decrease it. Impulsive desires or, as I have described them previously, those desires that just bubble up to our conscious mind, are a consequence of the questionable origin of our desires.

Overvold (1980: 108) describes a situation involving a man of modest means and his four sons. The man kills himself so that each of his sons will be able to attend expensive private colleges by using the proceeds from his large life insurance policy. Overvold argues that satisfying this apparently *self-sacrificial* desire does not make the man's life go better for him.

Heathwood (11) devises the case of Ellie and the music to be played at her 50th birthday party.[32] From her teens until a month before her party, she had a desire to have rock 'n roll played at her party. With a month to go and for the rest of the time up to and through her party, Ellie will prefer easy listening and will not be pleased if rock 'n roll is played at her party. A desire-satisfaction theory that counts past desires and how long the person had the desire will probably claim that we make Ellie's life go better by playing rock 'n roll at her party. Heathwood claims, rightly, that this is an implausible claim given the fact that Ellie's *changing* desire has now ensured that she will hate the rock and the roll.

Carson (2000: 72) describes a case in which he is thirsty and has a desire to drink from a nearby stream. There is a leak from a chemical factory upstream that he does not know about. Thus, satisfying this *ill-informed* desire

[32] Brandt (1979: 249-50) describes a similar case.

will kill Carson. This type of desire also seems problematic for a desire-satisfaction theory.

Schwartz (1982: 196) discusses the possibility of conditioning creating desires within us that may be contrary to our "true needs." Commercial advertising, religious training, political propaganda, and peer pressure are all mentioned by Schwartz as potential causes of *malconditioned* desires. Once again, the origin of some of our desires seems as though it may cause problems for a desire-centric theory.

Sumner (1996: 126) asks us to consider your possible desire to be remembered by your lover after you die. This desire, if it is satisfied at all, will necessarily be satisfied after your death. However, is it plausible to claim that this *posthumous* desire satisfaction actually makes your life go better for you? This is another potential problem created by our ability to desire anything at all.

Kraut (1994: 40-1) creates a scenario in which a man decides to punish himself for a crime he committed earlier in his life by taking a boring, arduous, insignificant job for several years after quitting his job that he loves. He considers it a moral necessity to satisfy this *desire not to be well off*. Satisfying this desire, by its own terms, seems to make one worse off, thereby creating another problem for desire satisfactionism.

After reading the letters of Keats and Van Gogh, Parfit (1984: 171) reports wanting it to be true that they knew how great their achievements were. This desire about the past—a *now-for-then* desire as Hare (1981: 101) calls

it—complicates matters as well. Could my welfare be affected by things that either did or did not obtain long before I was born?

Schwartz's (1982: 196) Bertha seeks above all else to minimize pain, and she is fully aware of the fact that going to the dentist in the near future would minimize her pain in the long run. Nevertheless, Bertha desires not to go to the dentist. Bertha's *irrational* desire provides another concern for the prospective desire-satisfaction theorist to sort out.

Heathwood (23) describes a case in which Father wants Son to get good grades, but only as a means to the end of what Father really wants for Son—"to be respected, to do worthwhile things, and to have a good life." Father's desire for Son to get good grades is an *extrinsic* desire (i.e., a desire that a state of affairs obtain only because of what it will lead to and not because the state of affairs is desired in itself). Is it plausible to claim that the satisfaction of desires that merely lead to things that are intrinsically desired makes one's life go better?

Finally, there is a group of various types of desires the satisfaction of which has been claimed not to make one's life go better, and perhaps make it go worse in some cases, due to the objects of the desires being *unworthy* of desire in some respect. First, one can have *malicious* desires. Charles Manson's desire to incite an all-out race war is a prime example. Second, one can have *tasteless*, *base*, or *poorly cultivated* desires. Preferring Justin Bieber, bestiality, or Birmingham, Alabama over available alternatives might be considered examples of these types of desires. Lastly, one could have pointless

desires. Rawls's (1999: 379-80) man who devotes himself to counting blades of grass or Kraut's (1994: 42) man who devotes himself to knocking down icicles are both examples of this type of desire.

This list of potentially problematic desires is meant to be exhaustive in order for us to completely demarcate the boundaries of the minefield we are about to traverse. On our journey we will step on some of these mines only to learn that they are duds. The rest we will avoid with the theory we will develop in the next three chapters.

CHAPTER THREE: UNIFIED THEORIES & FRANKFURTIAN FOUNDATIONS

Having chronicled all the potential pitfalls that await a desire-satisfaction theory of welfare, it is time to turn to an evaluation of the tools available to a desire satisfaction theorist that might be employed to fashion an adequate theory. This chapter will be devoted primarily to the tools provided by the work of Harry G. Frankfurt (b. 1929). Although Frankfurt has never addressed the question of personal welfare in print, the seeds for the theory I will construct and present over the course of the next few chapters can be found throughout his work. Frankfurt's works on free will, moral responsibility, rationality, and moral psychology formulated over the last 40 years are particularly helpful in this regard.

I. A UNIFIED THEORY OF WELFARE?

Before we turn to Frankfurt's ideas, a preliminary question about theories of welfare generally should be addressed. The question is a meta-axiological question concerning whether the right theory must include all sentient beings. The notion that this is what the right axiological theory should encompass is occasionally mentioned in passing in the literature,[1] but there does not appear to be any extended discussion of this issue.

The motivation for wanting a theory of welfare for all sentient beings is obvious. At least since the time John Stuart Mill (2006: 322) made the

[1] See, e.g., Kraut (1994: 47), Sumner (1996: 14), and Griffin (1986: 315 n.19).

following remark, a unified theory has been on the wish list of many an ethical theorist:

> It is better to be a human being dissatisfied than a pig satisfied; better to be Socrates dissatisfied than a fool satisfied. And if the fool, or the pig, are of a different opinion, it is because they only know this side of the question. The other party to the comparison knows both sides.

This is a controversial statement. Each time I teach Mill's *Utilitarianism*, there is always a subset of students who disagree with this claim. In fact, many utilitarians (e.g., Bentham) present a version of utilitarianism that would commit them to rejecting this claim. A unified theory would give us an answer to the question Mill raises.

Before we proceed further, the concept of a "unified theory" should be clarified. This concept could be defined in two ways. The first of these I will call a *strong unified theory*. A strong unified theory of personal welfare would measure welfare for all beings that have a welfare and would base welfare on the same attribute across every type of being. Hedonism might be an example of a strong unified theory if the version in question purported to cover all relevant beings and the measure of welfare across all of them was a function of the balance between pleasure and pain. If such a strong unified theory were the right theory, then, obviously, the theory of welfare (covering only persons) that I will put forward later would be false. Accordingly, I will offer some reasons in the rest of this section—in addition to the reasons I have already produced to doubt the truth of hedonism—to doubt the truth of such a theory.

The other possible kind of unified theory I will call a *weak unified theory*. A weak unified theory could allow for the welfare of each type (or even, perhaps, each instance) of being to be determined by different or unique attributes and would then specify a way to compare the welfare of any two beings—even beings of different kinds—in order to determine which life went better for the being that lived it. While the truth of a weak unified theory would be consistent with the truth of my theory, in this section I will give some reasons to doubt the commensurability of the welfare of different beings—which would mean no weak unified theory exists—and reasons to suppose that even a weak unified theory is radically epistemically inaccessible to us. And, at any rate, this discussion helps to frame and motivate the material from Frankfurt that follows it.

Returning to Mills's porcine idea, not only is there reason to *want* a theory that covers all sentient beings, there seems to be good reason to think that this is a good-making feature of a theory. If there were two competing theories, and one covered all sentient beings while the other covered some subset of sentient beings, it does seem reasonable to suppose the larger ambit of the former is a mark in its favor. However, to focus too much on this good-making feature of a theory may be to lose sight of the best good-making feature—that it be the *right* theory. If the right theory for, say, persons were only to include persons, then the problem is with our expectations of the theory and not the theory itself. Moreover, the idea that the right theory of welfare must include all sentient beings is a substantive claim that would need

supporting arguments. The claim that we have reason to *want* a unified theory is just that—a claim that we want such a theory and not a claim either that there is such a theory or that this is a requirement of the right theory. In fact, it seems that the best, and perhaps the only, way to argue that the right theory is a unified theory would be to put forward a theory, support the theory with arguments, and then take note of the fact that, as it turns out, this theory covers all sentient beings. Any attempt to argue that the right theory must be a unified theory using some sort of conceptual or analytic argument, without more, seems doomed from the start.

There is an additional reason to think that there is no strong unified theory that covers all sentient beings or that requiring one is really an attempt to stack the deck in favor of a particular theory. Which of these is the case turns on how one defines the term "sentient." In common usage, the term sentient means "having the power of perception by the senses; conscious" (Webster's Dictionary 1996: 1745). Without getting bogged down too deeply in complicated issues of philosophy of mind,[2] this definition of "sentient" opens the door to a potentially broad range of beings, which would seem to significantly reduce the prospect of finding a strong unified theory for all of them. Robots and other machines might fall into this definition. What makes a robot's life go better for it? Does this question make sense? Is this question

[2] "It is no accident that so many thinkers, both philosophers and scientists, have spoken of the 'mystery' of consciousness. It is not an exaggeration to say that the mystery of mind is, in essence, the mystery of consciousness" (Kim 2006: 224).

more like what makes a person's life go best for her, or more like what makes the life of Xerox copier go best for it? Is it like both questions? Is it in some third category such that it is like neither question? God and the Flying Spaghetti Monster would also presumably qualify as sentient beings under this definition. What makes their lives go better for them? Does *this* question make sense? If it does, how would one even begin to go about answering it? These issues point to the fundamental problem with defining "sentient" in this way. The problem is that by defining "sentient" as having the power of perception by the senses, this is the only attribute that all sentient beings can be assured of having, and almost all of these beings will have many of the additional attributes that it is possible for a being to have. If the power of perception by the senses is the only attribute that the beings at issue are assured of having, then a theory of welfare that encompasses all of these sentient beings will have to take one of two approaches. The first approach, a version of a strong unified theory, would be to make welfare a function of the only thing you can be sure every sentient being has—the power of perception by the senses. This option does not appear to hold much promise. In this option, it seems that welfare would increase by having *more* perceptions, *specific* perceptions, or *good* perceptions. Each of these options is so imperfect that explicit refutations are unnecessary. The second option, a version of a weak unified theory, would require an additional assumption in order to be possible—namely, that every sentient being *does* possess an attribute in addition to being sentient. If so, then each sentient being's welfare could be a function of this additional

attribute or this additional attribute and the power of perception by the senses. A detailed treatment of this very imperfect solution will be discussed below.

The other possible definition of "sentient" raises concerns of a different sort. Peter Singer (1989), in his famous article "All Animals Are Equal," described sentience as "the only defensible boundary of concern for the interests of others." In so doing, it seems as though he could not have had in mind the definition discussed above. Indeed, Singer acknowledges as much in a parenthetical that appears directly before the quote in the last sentence: "sentience (using the term as a convenient, if not strictly accurate, shorthand for the capacity to suffer or experience enjoyment or happiness)." For better or worse, Singer's "not strictly accurate" definition of "sentient" seems to have changed the definition, at least as the term is used in ethics, for the foreseeable future.[3] When the term "sentient" is defined in this manner, the concern for theories of welfare is obvious. For if a theorist, in developing a theory of welfare, is given the constraints that (1) the right theory must include all sentient beings and (2) sentient means the capacity to experience pleasure and pain, then it will be no surprise when the theorist produces a hedonistic theory of welfare. This is the concern noted earlier about stacking the deck in favor of a particular theory, in this case hedonism. And hedonism is almost surely the theory that would be produced given the two constraints above, as it is one of the two possible outcomes and is by far the more plausible of the two. This is because the only attribute we can be sure every sentient being, so defined, has

[3] See, e.g., Regan (2001: 71): "sentience (that is, the capacity to experience pleasure and pain)."

is the capacity to experience pleasure and pain (although, as I have argued in Chapter One, it is not a plausible theory for persons). The other possibility is only a possibility if we add in the additional assumption that was discussed at the end of the last paragraph, namely that every sentient being possesses some attribute in addition to being sentient. We could then have a weak unified theory that claimed the welfare of sentient beings was a function of this additional attribute alone or some combination of this additional attribute along with the capacity to experience pleasure and pain. As promised at the end of the last paragraph, the implausibility of this type of theory is what we will turn to next.

I have been arguing that the idea that there should be a strong unified theory for all sentient beings is a substantive claim that needs support and that there are reasons to think no such theory would be plausible.[4] This should not be confused with the claim that the lives of sentient beings, properly defined, *do not* go better or worse for the beings that live them. If this is correct, and if even a weak unified theory of welfare is neither possible nor plausible, then we may not be able to evaluate Mills's claim regarding humans and pigs that provided a great deal of the motivation for a unified theory in the first place. An illustration will be useful here in demonstrating the issue facing a weak unified theory. Let us suppose that the welfare of a pig is a function of

[4] Such a theory is implausible because, as I argued in Chapter One, hedonism is implausible (at least as applied to persons) and a theory of welfare for all sentient beings—when sentience is defined as pertaining to sensory perception—is even less plausible.

some attribute F that all pigs have. Although nothing important hinges on this, we can even suppose, perhaps plausibly, that attribute F is the capacity to feel pleasure and pain, which would then make a hedonistic theory of welfare the correct theory of welfare for pigs. We then find a happy pig, Babe, and calculate a welfare score of 100 for Babe, which, as it turns out, is quite a good score for a pig.[5] Next, let us suppose that the welfare of a person is a function of some attribute G that all persons have but that pigs do not have (which is what I will argue for in the coming pages). We then find a dissatisfied person, Arthur, and calculate a welfare score for him of 5, which is quite a low score for a person. We now want to know who had the better life—and thus, which life it would be better to have—Babe with a pig score of 100 or Arthur with a person score of 5. The proponent of a weak unified theory has a tall order on her hands if she is to answer this question. And notice that she *must* answer this question if she is truly offering a *unified* theory.[6] The task the unified theorist is faced with is devising a *pig-person welfare exchange rate*. In other words, she must specify an equation that allows the pig welfare score to be directly

---

[5] That'll do, pig. That'll do.

[6] Of course a theorist could maintain that there is a weak unified theory of welfare without actually specifying the theory. Such an attempt might rely on a hypothetical like the following. Imagine a super-duper great pig life and a terrible, horrible human life. Which one would you prefer? Most people, including me, would probably choose to live the pig life. However, this does not establish that there is a fact of the matter as to which life is better. As I have been explaining, I doubt that there is a fact of the matter. The reason, then, that I would choose the pig life over the human life is that I have some idea how a horrific human life would go and I would choose the good life of anything (or no life at all, for that matter) over the bad human life. This preference does not entail that it is *true* that the pig life is better than the human life.

compared to the person welfare score. For example, suppose the theorist claims that person welfare is multiplied by 2.5 in order to yield the equivalent pig welfare units. We would then know that Babe had a better life than Arthur by a score of 100 to 5 times 2.5 or 12.5. Such a theory would be wildly speculative at best. If there is a fact of the matter regarding which of these two lives is better, our lines cannot plumb these metaphysical and epistemological depths.

There is yet another consideration that makes the possibility of a successful weak unified theory unlikely. I have previously claimed that it makes sense to ask what life is like for the being that lives it and that it does not make sense to ask this question for other things (e.g., a Xerox machine). The problem with limiting the ambit of the former group to sentience, where sentience is defined as the capacity to experience pleasure and pain, is that, as explained previously, it will have a tendency to beg the welfare question in favor of hedonistic theories. Properly defining the boundary for a group whose welfare it makes sense to inquire about will illuminate another reason for us to be skeptical about the prospects for a unified theory. The proper boundary for welfare is all and only those beings for which there is something it is like to *be* that being (i.e., something it is like *for* that being or how it is for the being itself).[7] Following Thomas Nagel (1974: 436), we can call this the "subjective

[7] I am assuming that it does not make sense to inquire about the welfare of plants or inanimate objects. Things do not go better or worse *for* them. We may like one plant better than another, but that does not mean things are going better *for* that plant than they are for the other plant. And to the extent that such talk makes sense, I would argue that this is not talk about welfare,

character of experience." This presents a serious problem for even a weak unified theory of welfare if we add in an additional premise: In order to develop and adequately support a theory of welfare for a particular being (*i.e.*, how well the life goes *for* the being who lives that life), the theorist must know the subjective character of that being's experience. For if the theorist does not have this information, then is she not just throwing darts in the dark? She may hit the bull's-eye, but she would never have reason to believe this without turning on the lights (supplying her with the required subjective character of experience), nor would we as spectators (the evaluators of her theory) have reason to believe this. This is because, as Nagel so persuasively argues, the subjective character of experience for some (or all?) creatures is radically inaccessible to us. Supposing that there is something that it is like to be a bat, Nagel asks what it is like for a bat to be a bat. In other words, what is the phenomenal character of a bat's experience for a bat from navigating the world primarily through the use of the bat's incredibly discriminating sonar system? One could know all the physical facts about a bat for any period of time and still not have the faintest idea what it is like for a bat to be a bat. As Nagel (1974: 439) says, if he tries to imagine this, "I am restricted to the resources of my own mind, and those resources are inadequate to the task. I cannot perform it either by imagining additions to my present experience, or by imagining segments gradually subtracted from it, or by imagining some combination of additions, subtractions, and modifications." So while there is

---

but about some other concept. I will have to leave these claims undefended here.

something that it is like for a bat to be a bat, Nagel (1974: 441) claims that our lines cannot plumb these depths either:

> Reflection on what it is like to be a bat seems to lead us, therefore, to the conclusion that there are facts that do not consist in the truth of propositions expressible in human language. We can be compelled to recognize the existence of such facts without being able to state or comprehend them.

This may be the best reason to think that even a weak unified theory of welfare is beyond our reach at best and perhaps does not exist at all.

There is one final consideration that may cast doubt upon the accessibility or even existence of a unified theory and will also serve as a smooth transition to our next topic. This issue presents itself if we take a closer look at the relationship, if any, between the welfare of nonhuman animals—let us just use pigs for simplicity's sake—and the welfare of persons. Now as I have stated previously, every theory of welfare—in order to be a theory of welfare at all—must specify *something* that makes a life go better for the being who lives that life. Although I have argued against the existence of a unified theory for all sentient beings, perhaps there is a unified theory for nonhuman animals and persons. While I do not think this is the case either, a cursory examination of this possibility should prove useful. If there were a unified theory for pigs and persons, the three most plausible options are an objective-list theory, a hedonistic theory, or a desire-satisfaction theory. An objective-list theory is a nonstarter. First, bearing in mind the lengthy argument in Chapter One against the possibility that an objective-list theory is the right theory for persons, it follows that an objective-list theory cannot be

the correct theory for persons *and* pigs. Second, a weak unified objective-list theory would be faced with the daunting pig-person welfare exchange rate problem described above, and a strong unified objective-list theory would have to specify the same list for both pigs and persons; neither of these approaches seems promising. A hedonistic theory for both pigs and persons, while more plausible than objective-list theory, also appears to be false. Although a hedonistic theory for pigs is a good guess (what is it like to be a pig?), I argued against this being the correct theory for persons in Chapter One as well. Again, if it is not correct with respect to persons, it cannot be correct for persons and pigs.

This brings us to the possibility that a desire-satisfaction theory is the correct theory for both pigs and persons. Could desire satisfactionism be the correct theory for pigs? Although I do not want to commit myself to any definitive answer here because nothing hinges on this in terms of my larger project, I suspect there is at least one good reason to reject this possibility. It seems reasonable to suppose that any adequate theory for animals must countenance the fact that pleasure makes an animal's life go better and pain makes it go worse. We were able to accommodate this data point with respect to persons by appealing to the Motivational Theory of Pleasure (MTP). It is not clear to me that this strategy will be successful when it comes to nonhuman animals. With respect to identical sensations, could a pig desire it over time, thus making it pleasurable, and not desire it at another time, thus making it

painful?[8] Whether MTP works for nonhuman animal welfare, or could be modified to do so, is a question I cannot answer here.

If an unrestricted desire-satisfaction theory will not work for pigs, what about a restricted theory? This is where this exercise gets interesting, as this *is* what I will claim is the correct theory for persons. What I would like to examine here is not the answer to the following question, but the implications of the answer to the question: Are the welfare-producing desires restricted in the same way for pigs as they are for people? In other words, will the restricted desire-satisfaction theory that I intend to claim is the correct theory of welfare for persons also work for pigs? The answer is no. If any restricted desire-satisfaction theory could work for pig welfare, it is not the one I will offer in the coming pages. However, let us suppose the answer was yes; there is a restricted desire-satisfaction theory that could be used to determine the welfare of *any* animal—human or nonhuman, person or pig. Leaving aside any speculation of what this theory might look like, there is an interesting implication. If this were right, then the correct theory of welfare for persons not only would not, but *could not*, include any feature that makes humans distinct from other animals. This outcome would be a rather large surprise and is yet another reason to suppose that a strong unified theory of welfare is not possible. This is because the difference between human animals and

[8] The thought that I cannot adequately explore here is that perhaps some type of reflective attitude might be required in order to want a sensation at one time and not want the exact same sensation at another time. If something like this were true, it would make MTP inapplicable to most nonhuman animals and, in turn, would make a desire-satisfaction theory incorrect for them.

nonhuman animals is not a trivial characteristic that *would* be surprising to find was related to human welfare—like, say, if we were the only animals that laid eggs or had horns. What does set persons apart is our capacity for self-consciousness, which seems to give rise to most, if not all, of the aspects of our mental lives that we value the most: free will, rationality, highly refined language skills, etc. A strong unified theory of welfare for all animals would mean that none of these things could impact our welfare. While this may ultimately be true, it would be at odds with most people's intuitions about welfare and may be a sign that the theory claiming this needed to be reexamined.

II. What Is a "Person"?

If one does not intend to offer a unified theory of welfare, then it is a good idea to be very clear about the sort of being for which one does intend to offer a theory of welfare. As I will be offering a theory of *personal* welfare, I will attempt to clarify what the term "person" means in this context and why clarity on this issue is so important.

What does it mean to be a *person*? It seems clear that a person must be a being of some sort, but what sort? To answer this will be to specify the property or properties that a being must have in order to be a person. In other words, persons are all and only those beings who possess property F.[9] What property, then, should property F be? In common parlance, property F is

[9] For a detailed discussion of various ways "person" has been defined, see Loren E. Lonsky's (2001: 1293) "person, concept of" in the *Encyclopedia of Ethics*.

assumed to be equivalent to human being so that persons are all and only human beings. This assumption is even made by some philosophers (see, e.g., Wertheimer 1971: 69). Although defining personhood in this way is a natural tendency and quite harmless on many occasions, it may be problematic in many philosophical contexts.

Often when philosophers use the word "person", it connotes a sense of elevated moral standing such that persons are often claimed to have rights of some sort (e.g., the right to life, the right to bodily integrity, etc.). One problem with defining persons as human beings is that it will not include many possible or actual beings that would seem to possess the same rights. This problem stems from the fact that to be a human being is just to be a being with a specific sequence of DNA. Is a specific sequence of DNA a morally relevant difference such that those with it have rights and those without it do not? The answer would seem to be no. If a race of beings were discovered that were identical to humans in every way except that they had XYZ instead of DNA, would it be morally justifiable to rape, torture, and kill them as we pleased? Intelligent life forms on other planets, if they exist, might be interested in our answer to this, given the thought that what is good for the goose is good for the gander.

The other problem with equating persons and human beings pertains specifically to the question of welfare and also relates back to the point made at the end of the last section. This problem concerns the fact that the development of a human being is a process involving continuous and gradual

change. This ground has been covered thoroughly in the abortion debate (see, e.g., Thomson 1971: 47-48). The problems caused by this fact in the abortion debate raise analogous issues in the welfare debate. This is because the thing formed is biologically human very soon after, or perhaps at, conception. (This makes sense; what else, biologically, could it be?) However, we do not have a problem yet from an axiological perspective. This is because, at least up until about the 25th week of development, there is nothing it is like to be one of these things (*i.e.*, there is no subjective character of experience), as the central nervous system is not yet functioning. From the standpoint of welfare prior to this point and excluding any potential future welfare, this biologically human entity is nothing more than a glorified tumor until, at the very least, the start of brain function. Once there is something that it is like to be this thing, though, it makes sense to ask about its current welfare. Leaving aside the underinclusiveness issue discussed in the last paragraph, shall we say that persons are humans who can be said to have the subjective character of experience? This definition would also be an unfortunate one. If our theory of welfare for persons must include you, me, and fetuses, what shall we specify that makes our lives go better and that also makes a fetus's life go better? Perhaps the only option, and at least the most plausible of the very few options, will be our previously discredited hedonistic theory of welfare. The options do not get much better, either, until well after birth because we do not develop the mental features that separate us from other animals until long after birth.[10] If

[10] This can be seen quite nicely if one observes the similarities in how adults

a theory of welfare is to include both adults and fetuses/babies anywhere in this stage of development (leaving aside questions of potential future welfare), then we must choose from features that we share with other animals and we must exclude any features that arise from our capacity for self-consciousness.

Michael Tooley notes this same fact, at least as it relates to fetuses and babies, in pointing out the difficulty in developing a position on abortion that allows for abortions but does not allow infanticide (Tooley 1972: 37-38). In formulating his own position on abortion and infanticide, Tooley (1972: 40) treats the concept of a person as a purely moral concept that is synonymous with a being that "has a serious right to life." Tooley (1972: 44) goes on to argue that the necessary (and possibly also sufficient) properties a being must have to be a person (i.e., a being with a serious right to life) are that "it possesses the concept of a self as a continuing subject of experiences and other mental states, and believes that it is itself such a continuing entity." Tooley (1972: 44) calls this the "self-consciousness requirement." Here, finally, we are getting much closer to something robust enough to work with from an axiological perspective if one is to formulate a theory of welfare that can give at least some weight to the feature that separates us from the rest of the animal kingdom. Although the definition of personhood I intend to use is slightly different from Tooley's definition, for reasons that should become clear in the next section, there is a great deal of overlap between his project and this one. For one, the properties he requires for personhood are necessary preconditions

---

relate to and teach both puppies and babies. There is little difference in either the actions of the adults or the actions of the puppies and babies.

for the properties that I will claim constitute personhood. Moreover, although Tooley never says as much, the rationale for his self-consciousness requirement seems to be one that is deeply aligned with many of the arguments in the previous chapters. The rationale seems to be that you do not have a serious right to life (i.e., you are not a person) if you are incapable of *caring* that you are being deprived of that right.

III. Why Use Frankfurt's Definition of Personhood?

As mentioned at the beginning of this chapter, I intend to adapt many of Frankfurt's concepts and arguments for use in a new theory of personal welfare. While it is certainly the goal to clarify why Frankfurt's work is being used during the course of explaining it and laying out the new theory, there is good cause to touch briefly upon some of those reasons at the outset in order to evaluate them in relation to the bigger picture when they are presented later in this chapter and the next.

The central concept of Frankfurt's that I intend to make use of is his definition of personhood. According to Frankfurt, persons are all and only those beings who possess property F where property F is to be found in the *structure of the will*. As will be explained in great detail in the next section, a *person* must have a certain type of desire *about* her desires, which Frankfurt calls *second-order volitions*. The basic idea is that persons must have an evaluative attitude toward themselves as agents. This attribute that gives rise to personhood is made possible by the reflective component contained in Tooley's self-consciousness requirement, but it goes beyond Tooley's definition

to capture the attributes that are most important to ourselves. By locating personhood within the structure of the will, being a person becomes essentially tied to several salient features of beings like us.

First, being a person brings in the capacity for free will. If, as is commonly supposed, persons are capable of enjoying free will and most, if not all, other animals are incapable of free will, then this is a fact that needs to be explained and not just asserted. Frankfurt's account explains free will in a very straightforward manner by making use of the essential feature of persons—second-order volitions. Freedom of the will, according to Frankfurt, is achieved when a person secures the conformity of his will to his second-order volitions. Frankfurt also manages to explain free will coherently without resorting to any miraculous (by definition) absence-of-causal-determination claims—no small feat. Moreover, in the context of axiology, it should not be surprising to find that the value of a life is tied to notions of free will. Indeed, it seems as though the capacity for free will is not only an *essential* feature of persons, but a *valuable* feature as well. Frankfurt's account of persons and free will can explain this fact in a way that none of the other major personal welfare theories can.[11]

Second, relating personal welfare to Frankfurt's conception of personhood allows personal welfare to be linked with personal identity. To the extent that a person is defined by his will, then the theory of personal welfare I

[11] In Chapter Five, I will discuss how other theories will most likely make free will appear to be an ad hoc addition to the theory, or will have to assign free will an arbitrary value, or both.

am developing will be bound up with a person's essential nature or, if you like, his identity as an agent. This relationship also seems correct, at least from a common-sense standpoint, as it would be strange to find that a person's welfare was completely unrelated to his essential nature.

Third, Frankfurt's conception of personhood yields an excellent account of what it is to *care* about something. This is obviously critical to a theory of personal welfare, as the earlier discussion of the Principle Concerning Caring (PCC) demonstrated. According to Frankfurt, caring is essentially a volitional activity that consists in having, and identifying with, second- (or higher-) order volitions. In fact, the formation of the will is primarily an exercise in coming to care about certain things because the person wants the desires pertaining to these things sustained. Accordingly, what it means for something to be important to a person is just that the person cares about that thing. And it is through the act of caring that a person is provided with stable motivational structures that guide and limit his conduct. In other words, caring is the characteristic of persons that allows us to be involved in our own lives (i.e., not to care about anything is to be uninvolved in one's own life).[12] Should we be surprised, then, to learn that a person who does not care about *anything* has a life that does not go better or worse for him than no life at all?

Finally, Frankfurt's conception of personhood as being comprised of a certain volitional structure allows for a compelling account of the active-passive distinction. This is important from an axiological perspective if, as is

[12] As will be explained in the paradox of welfare section at the end of Chapter Five, caring about something makes it about our lives.

commonly supposed, there is some value in being active. Frankfurt (1988: viii-ix) frames this issue beautifully in the preface of his book, *The Importance of What We Care About*:

> In the seventeenth century, mechanism became established as the dominant worldview of our culture. It has since that time come to seem obvious that either references to final causes are entirely illicit or they are no more than convenient ways of speaking designed to avoid clumsier (albeit more strictly accurate) formulations in terms of efficient causation. In the eighteenth century, the notion of an efficient cause was itself eviscerated by a devastating critique of the idea of inherent power. These compelling philosophical developments have made it difficult to give a good account of the difference between being active and being passive. For if things are understood as having neither purposes nor powers, in what way is it possible to comprehend them as being active at all? Nonetheless, the role of the active-passive distinction in human life is pervasive and deep. The difference between passivity and activity is at the heart of the fact that we exist as selves and agents and not merely as locales in which certain events happen to occur.

Precisely how this works will be explained in detail later in this chapter along with the three topics that preceded it.

However, before turning to the detailed discussion of Frankfurt's ideas that I intend to adapt for this project, one final note regarding the use of Frankfurt's conception of personhood bears mentioning. With regard to the possibility of a unified theory of welfare for all sentient beings discussed at the beginning of this chapter, one of the recurring problems was that none of the plausible definitions of "sentient" yielded anything robust enough on which to base a theory of welfare. That is obviously not the case here. In Frankfurt's conception of personhood, we can be assured not only of the presence of desires, but also of a special class of desires resulting from our self-conscious,

reflexive capacity. This will prove to be an excellent foundation for our theory of personal welfare.

IV. Frankfurtian Persons

Frankfurt first presented his definition of personhood in his seminal 1971 article *Freedom of the Will and the Concept of a Person*. His account begins with desires that humans share with nonhuman animals and then builds to include the elements that are essential to personhood.

Before we proceed to Frankfurt's discussion, a note about his definitions of "person" and "care" are in order. In common usage, "person" often simply means "human being" or "Homo sapien" and "care" often means "desire." Frankfurt's definitions of each of these terms are more restrictive, and in that sense, are stipulative. However, in another sense, I do not think Frankfurt thinks of these definitions as being stipulative. So Frankfurt might claim that in formulating these definitions, he is trying to capture what, upon reflection, we are really trying to convey when we use these terms carefully.[13] In any event, both of these terms are central to the rest of the project, and paying close attention to what Frankfurt means by each will aid in understanding the coming chapters.

---

[13] This can be seen in a comment Frankfurt (1988: 12) makes about his interest in the term "person":

> In those senses of the word which are of greater philosophical interest, however, the criteria for being a person do not serve primarily to distinguish the members of our own species from the members of other species. Rather, they are designed to capture those attributes which are the subject of our most humane concern with ourselves and the source of what we regard as most important and most problematical in our lives.

At least in terms of those desires that we share with most nonhuman animals, Frankfurt (1988: 12) defines first-order desires as "simply desires to do or not to do one thing or another." First-order desires are expressed in the form "A wants to X" where "to X" refers to a possible action or inaction (Frankfurt 1988: 13). Frankfurt then makes the distinction between effective first-order desires and non-effective first-order desires (or simply first-order desires). He does this because, as he rightly notes, there are a multitude of reasons why a first-order desire may not result in an action or may have nothing to do with motivating the action that is taken. For example, the person may be unaware of the desire, the person may want to do something else more, the person may also want to refrain from taking the action in question, the person may take the action in question but be motivated to do so by an entirely different desire, etc. *Effective desires*, then, are a subset of first-order desires that move "(or will or would move) a person all the way to action" (Frankfurt 1988: 14). According to Frankfurt, an agent's will is identical to effective first-order desires.

However, this is not the end of the story for the agent's will. Second-order desires also pay a critical part. If first-order desires are desires expressed in the form of "A wants to X" where "to X" is an action or inaction, then second-order desires are desires expressed in the form "A wants to X" where "to X" refers to a first-order desire. In other words, a second-order desire takes the form of "A wants to want to X" (Frankfurt 1988: 15). Frankfurt discusses two varieties of second-order desires. While perhaps empirically scarce, the first

type of second-order desire discussed is important in that it demonstrates that having a second-order desire for a first-order desire does *not* entail actually having that first-order desire.

Frankfurt's example involves a psychotherapist treating drug addicts. The therapist believes that he would be better able to help his patients if he fully understood what it is like for his patients to desire the drug they are addicted to. Because of this, he is led to want to have a desire for the drug. In other words, he has a second-order desire for the desire to take the drug. However, he has no desire to actually take the drug and may, in fact, have a strong desire *not* to take the drug. "And insofar as he now wants only to *want* to take it, and not to *take* it, there is nothing in what he now wants that would be satisfied by the drug itself" (Frankfurt 1988: 15). Accordingly, having a second-order desire for a first-order desire does not entail actually having that first-order desire. Frankfurt says that someone who *only* wants to want a certain desire "stands at the margin of preciosity, and the fact that he wants to want to X is not pertinent to the identification of his will" (Frankfurt 1988: 15).

The second variety of second-order desire described by Frankfurt is critical for determining the structure of the will. This variety of desire described by "A wants to want to X" indicates what A wants his will, or his effective first-order desire, to be. In these cases, "A wants to want to X" "means that A wants the desire to X to be the desire that moves him effectively to act" (Frankfurt 1988: 15). If A wants this first-order desire to be effective (i.e., to provide the motive in what he actually does), then it does entail that A already

has the first-order desire to X. To claim otherwise, one would have to claim that it is true both that A wants the desire to X to be his will and that he does not want to X. Frankfurt's assessment that this is incoherent seems undeniable. And for purposes of a desire-satisfaction theory of welfare, it should also be made explicit that when A has a desire that a certain first-order desire be effective (i.e., his will), then it entails not only that A wants the desire in question to be the desire that moves him effectively to act, but also that A wants the desire in question to be satisfied. For just as it would be incoherent to claim that A wants the desire to X to be his will and that he does not want to X, it would be equally incoherent to claim that A wants the desire to X to move him to effectively act and that A wants the desire to X to be frustrated.

This second variety of second-order desires, desires about those desires that the agent wants to become her will, Frankfurt (1988: 16) calls *second-order volitions* and are essential to being a *person*. Frankfurt (1988: 16) contrasts this with the term *wanton*, which he defines as agents that have first-order desires—and perhaps second-order desires—but who do not have second-order volitions. Another way of stating the essential feature of a person is that he cares about his will; the essential feature, then, of a wanton will be that he does *not* care about his will. The class of wantons includes most, if not all, nonhuman animals, children,[14] and some adult human beings.

[14] The existence of second-order volitions in a particular agent is, of course, an empirical question. Children will all be in the class of wantons up until they develop the capacity for reflection. At that point they *can* become critically

A wanton has first-order desires, but no second-order volitions, and thus, he can be said not to care about his will. So a wanton's desires "move him to do certain things, without it being true of him either that he wants to be moved by those desires or that he prefers to be moved by other desires" (Frankfurt 1988: 16). In the case of conflicting first-order desires, it is not the case that the wanton is neutral in the conflict because he finds both desires equally acceptable. In failing to care about his will through the formation of second-order volitions, "it is true neither that he prefers one [desire] to the other nor that he prefers not to take sides" (Frankfurt 1988: 18). Furthermore, his failure to care about his will in cases of conflicting first-order desires is not due to his inability to find a compelling reason to prefer one over the other. Simply put, the wanton's failure to care about his will is due to one of two reasons. The first possibility is that the wanton lacks the capacity for reflection, such as might be the case with a young child or would be the case with a squirrel. The second possibility would be that the wanton possesses a "mindless indifference to the enterprise of evaluating his own desires and motives" (Frankfurt 1988: 19).[15] Accordingly, the wanton, or the person acting in a wantonly fashion in a particular situation, by definition, will "pursue

---

aware of their own will, and some of them will form second-order volitions and become persons, at least with respect to their actual second-order volitions.

[15] Although Frankfurt does not specifically address the question, I do not think occurrent higher-order volitions are necessary to avoid being a wanton with respect to any particular act or desire. As always having occurrent higher-order volitions could be exhausting, dispositional volitions should suffice. As long as the course of action has been properly reflected upon at some point, the resulting dispositional desires should be sufficient to address Frankfurt's theoretical concerns.

whatever course of action he is most strongly inclined to pursue, but he does not care which of his inclinations is the strongest" (Frankfurt 1988: 17). However, while a wanton either cannot or does not deliberate (i.e., perform a reflexive action that the mind does to itself), this does not mean that a wanton cannot act *intelligently*. This can be clearly seen by examining the case of some nonhuman animals. Although they lack any reflexive capacity, they do act intelligently quite frequently. There is no reason to believe that a wanton would not act with similar intelligence to satisfy his desires (Frankfurt 2006: 14).

This distinction between wantons and persons can be illustrated by a case that is hard to make sense of without utilizing these concepts relating to the will. Consider the cases of Dexter and Dahmer. Both have first-order desires to kill other human beings. Both also have first-order desires not to kill other human beings. Furthermore, both Dexter and Dahmer succumb to their desires to kill from time to time, such that an observer could not tell the difference between the actions of these two men. But here the similarities stop, for Dexter is a person and Dahmer is a wanton. What makes Dexter a person, if only in this situation, is that he has a second-order volition with regard to these competing first-order desires. Dexter wants the first-order desire not to kill to be his will, to be effective, to provide the motivation for what he does when he does act. However, Dexter is overtaken by his desire to kill from time to time—his dark passenger—despite his constant struggle against it. When this happens he is violated, against his will, by his own desire to kill. This is

because Dexter *cares* which of his conflicting first-order desires wins out in the end. This caring comes about through his second-order volition whereby Dexter *identifies* himself with his first-order desire not to kill and withdraws himself from the competing desire. It is through this process that we can make sense of the idea of Dexter's being moved by a force that is not his own. When he kills, which is directly contradictory to his second-order volition, he is moved by a force that is not *his own* and against his own will.

The wanton, Dahmer, on the other hand, does not care about his will. When he acts, it merely reflects the economy of his first-order desires. The strongest desire will win out, and it makes no difference *to Dahmer* which desire does so. He cannot win this battle, just as he cannot lose it. When Dexter acts he is moved by the will he wants or by the will he does not want. When Dahmer acts, it is neither.

V. What Is Wrong with Mere First-Order Desires?

Many concepts were introduced in the last section that will require a full explanation as well as a justification for their being used in the context of personal welfare. Most of them will be discussed in the remainder of this chapter, while a few will be left for the next chapter. The topic for this section, however, is first-order desires. As the last section made clear, desires are essential for personhood. Nevertheless, first-order desires alone are not sufficient. So what is the problem with first-order desires for purposes of both personhood and welfare?

The problems with first-order desires[16] can be broken down into two major categories. The first problem stems from the *origin* of first-order desires. While humans may be the only beings capable of higher-order desires,[17] humans share the capacity for having first-order desires with a large segment of the animal kingdom. And this shared capacity for first-order desires seems to operate in much the same way for both human and nonhuman animals. Desires just occur within us after being caused by a variety of hereditary and environmental factors. In most nonhuman animals, this means that their actions simply reflect the economy of their first-order desires, with, other things being equal, the most intense desire at the time winning out over the others.

In persons, however, this plays out differently and, as a result, helps to highlight the second problem with first-order desires—the *role* these desires play in our mental lives. Instead of the unfettered sovereign reign of first-order desires found in most nonhuman animals, persons simply *find* these desires within themselves. First-order desires merely bubble up to the surface—in response to environmental factors—out of the murky, evolutionary stew that is

---

[16] "First-order desires," as used in the rest of this section, means neither first-order desires prior to any related higher-order desires occurring nor first-order desires that were generated by any higher-order desires.

[17] It should be noted that nothing of consequence hinges on this claim. If, say, dolphins and great apes have higher-order desires as a result of developing self-consciousness, then they too would be persons. Moreover, there are no obvious evolutionary obstacles preventing other species from coming to have the capacity for self-consciousness and higher-order desires in the future.

our minds.[18] Accordingly, they are just "psychic raw material," and the *person* must decide if, and how, he is to incorporate them into the structure of his will (Frankfurt 1999: 137). From this we can see that first-order desires are not volitional, but rather are "impulsive or sentimental," as both their origin and role would suggest (Frankfurt 1999: 99).

Equipped with this description, it is much easier to make sense of the possibility that these are desires *that belong to no one*. In other words, these are not desires that are attributable to the *person* since they are ones that the person merely finds occurring within her body. This does sound odd, and it is opposed by a simple and straightforward argument from Terence Penelhum (1971: 674):

> Premise 1: Every desire must belong to someone.
> Premise 2: A desire that occurs within a person cannot belong to anyone else.
> Conclusion: Therefore, every desire belongs to the person whose body it occurs within.

There is a literal sense in which this is quite obviously true. In much the same way as Descartes (1996: 17) claimed that the fact that there was thinking led him to the conclusion that there had to be *something* doing that thinking, Penelhum concludes that desires require a desirer, and the desirer is the person whose body the desires occur within. However, this argument glosses over an important distinction between persons and nonhuman animals. Penelhum's argument works quite well with most nonhuman animals, as for

[18] There is also the possibility of desires that do not bubble up to the surface and, accordingly, are subconscious desires. Also, as is evident from the claim in the text, I am assuming that innatism is false, although nothing of importance hinges on this claim.

them there is no interesting or relevant distinction to be made concerning their desires. The strongest desire, other things being equal, will be translated into action at any given moment. Persons, on the other hand, can reflect upon desires they find occurring within their bodies and decide whether a given desire is one they want to act upon or one they want to distance themselves from so that the desire will not be acted upon. In this way, it is much less jarring to say that, while a desire *occurs within* a person's body, it is not a desire that *belongs to* the person. In fact, this type of claim seems trivial and commonplace in an analogous context pertaining to persons. In one sense, all bodily movements must be movements occurring within a body. This statement, however, overlooks an important distinction in that not every movement of a body is a movement performed by the person. There are, after all, a large class of movements that are not performed due to any conscious interest on the part of the person. For example, my body may perform a reflex movement such as when a doctor taps my knee with a hammer or when I blink as the result of a loud noise. My body also performs a large number of *involuntary* movements such as my heart beating. It is also possible to have seizures or spasms, such that there are definitely actions being performed that are not the person's doing. Frankfurt (1988: 61) sums up this idea nicely: "A person is no more to be identified with everything that goes on in his mind, in other words, than he is to be identified with everything that goes on in his body."

If persons are not to be identified with all of the desires that occur within their bodies, then this fact obviously has further implications concerning the role these first-order desires play in our psychic makeup. While there will be much more on *caring* later in the chapter, it is worth pointing out here that the fact that a person is not to be identified with all of the desires occurring within her body seems to entail that some desires are ones the person does not care about. An example may help make this point clear. Suppose I find myself wanting a new gadget, and I have narrowed it down to the iPhone or the iPad, of which I can only afford one. After examining the relative strength of my desire for the iPhone and my desire for the iPad, I discover that my desire for the iPhone is stronger and, thus, I would prefer the iPhone. Does this fact entail that I care about my desire for the iPhone? No. "The fact that a person wants one thing more than he wants another does not entail that he cares about it more, therefore, because it does not entail that he cares about it at all" (Frankfurt 1999: 157). This is the proper conclusion to draw given that a person's stance on desires he finds within himself may be anything from caring very deeply about them to being completely appalled that they are occurring to anything in between (including not paying attention to them at all other than, perhaps, noting their presence).

Finally, having characterized first-order desires as psychic raw material that may belong to no one and that the person they occur within may not care about, it would be easy to be almost completely dismissive of these desires. This would be a mistake, as these desires can be quite powerful and persistent.

Take, for instance, the example of teenage boys and the effect that their hormones have on their desires. How can we properly characterize the role these types of desires, or *passions* as Frankfurt sometimes refers to them, play in our motivational structures without minimizing the intensity that they can have? Frankfurt (1999: 137) does a perfect job:

> However imposing or intense the motivational *power* that the passions mobilize may be, the passions have no inherent motivational *authority*. In fact, the passions do not really make any *claim* upon us at all. Considered strictly in themselves, apart from whatever additional impetus or facilitation we ourselves may provide by acceding to them, their effectiveness in moving us is entirely a matter of *sheer brute force*. There is nothing in them other than the magnitude of this force that requires us, or even that encourages us, to act as they command.[19]

---

[19] Nietzsche (2005: 171-72) is also helpful on this topic and, in his own way, is in apparent agreement with Frankfurt—as we will see later—on how desires we find within ourselves can ultimately be beneficial:

> All passions go through a phase where they are just a disaster, where they drag their victim down with the weight of their stupidity – and a later, much later phase where they marry themselves to spirit, where they 'spiritualize' themselves. People used to fight against the passions because the passions were so stupid: people conspired to destroy them, – all the old moral monsters are unanimous on that score: '*il faut tuer les passions*'. The most famous formula for this is in the New Testament, in that Sermon on the Mount, where, incidentally, things are certainly not viewed *from a higher perspective*. When it comes to sexuality, for instance, it says: 'if your eye offends you, pluck it out': fortunately, Christians do not follow this rule. Nowadays, to *destroy* the passions and desires just to guard against their stupidity and its unpleasant consequences strikes us as itself a particularly acute form of stupidity. We have stopped admiring dentists who *pluck out* people's teeth just to rid of the pain . . . But it is reasonable to admit that the idea of '*spiritualizing* the passions' could never have arisen on the soil where Christianity grew. It is well known that the first church even fought *against* the 'intelligent' for the sake of the 'poor in spirit': how could we expect it to have waged an intelligent war on the passions? – The church combats the passions by cutting them off in every sense: its technique, its

From this discussion of the origin of our first-order desires and their role in our motivational structures, it becomes clear that it would be dogmatic to claim that the satisfaction or frustration of every first-order desire a person has must be relevant to assessing his well-being. However, this is the claim that William James (1948: 73) seems to be making when he writes, "Take any demand, however slight, which any creature, however weak, may make. Ought it not, for its own sole sake, to be satisfied? . . . Any desire is imperative to the extent of its amount; it makes itself valid by the fact that it exists at all." This claim, as we have repeatedly seen, is far too indiscriminate. Not only does it overlook important distinctions in both the type of creature that has the desire and the type of the desire itself, it also ignores the obvious fact that not all unsatisfied desires are frustrated. A person may simply decide to give up a desire by no longer wanting the object in question. All of this should not be construed, of course, to endorse the idea that desires are not relevant to personal welfare. The claim that not every desire impacts personal welfare does not entail the claim that no desire impacts personal welfare. As has been said before, the trick, and the point of this project, is to properly discriminate between desires that impact welfare and those that do not.

> 'cure', is *castration*. It never asks: 'how can a desire be spiritualized, beautified, deified?' – It has always laid the weight of its discipline on eradication (of sensuality, of pride, of greed, of the thirst to dominate and exact revenge). – But attacking the roots of the passions means attacking the root of life: the practices of the church are *hostile to life* . . .

VI. Reflective Capacity & The Will

In previous chapters, I have attempted to make the case that some form of desire satisfactionism must be the correct theory of personal welfare because (1) some lives *do* go better than others, and (2) none of the competing theories are plausible. In the previous section, I have attempted to demonstrate that first-order desires (having characterized them as psychic raw material that may not even belong to the *person* in question), without more, cannot be what makes the life of a person better or worse. By process of elimination, then, higher-order desires, or some subset thereof, must be what drive personal welfare. Nevertheless, in the remainder of this chapter I intend to make the positive case for this outcome rather than simply relying on a relatively unsatisfying process of elimination.

> It happens that the stage sets collapse. Rising, streetcar, four hours in the office or the factory, meal, streetcar, four hours of work, meal, sleep, and Monday Tuesday Wednesday Thursday Friday and Saturday according to the same rhythm—this path is easily followed most of the time. But one day the "why" arises and everything begins in that weariness tinged with amazement. "Begins"—this is important. Weariness comes at the end of the acts of a mechanical life, but at the same time it inaugurates the impulse of consciousness. It awakens consciousness and provokes what follows. What follows is the gradual return into the chain or it is the definitive awakening. . . . For everything begins with consciousness and nothing is worth anything except through it. (Camus 1996: 358-59)

For better or worse, rational human beings seem to be the only sort of creature for whom this type of occurrence is possible. Other creatures are only capable of a purely mechanical life, meaning, roughly, their instincts combined with their environment strictly determine what they will do at any given time.

For them, there can be no "awakening" of the type Camus discusses. And while Camus's discussion of the awakening is a bit overly dramatic, it does serve to nicely highlight the point I wish to make here. The "awakening" and the "consciousness" of which Camus speaks are, of course, both referring to a person's coming to use his capacity for self-consciousness. Self-consciousness, Camus appears to claim, is at least necessary for anything to have value. This is obviously a very bold claim, but not an altogether implausible one. It seems relatively straightforward that there could be no *meaningful discussion* about value in a system that contained no instances of self-consciousness. If that is true, then there does not appear to be any non-arbitrary way to assess, determine, or assign value in such a system. Nothing of significance, however, rests on those claims here. I intend to support the much more modest claim that self-consciousness is a necessary condition for personal welfare.

So *why* should we care about self-consciousness? The answer to this question is best answered by examining *how* self-consciousness impacts our lives. To do this, a particularly difficult thought experiment should prove useful. Stop reading for a moment and try to imagine what *your experience of yourself* would be like without the capacity for self-consciousness.

Welcome back.

Let me begin with an apology for asking you to do the impossible. Without the capacity for self-consciousness, you would not, because you could not, have any experience of yourself. In this way, you would be no different from the nonhuman animals Frankfurt (2004: 18) describes:

> They are moved into action by impulse or by inclination, simply as it comes, without the mediation of any reflective consideration or criticism of their own motives. Insofar as they lack the capacity to form attitudes toward themselves, there is for them no possibility either of self-acceptance or of mobilizing an inner resistance to being what they are. They can neither identify with the forces that move them nor distance themselves from those forces. They are structurally incapable of such interventions in their own lives.

This would obviously be a significant and—I am assuming—unwelcome change in our mental lives. And while I will be discussing some of the features that we would lose in the coming pages, it would do to reflect on just how different we would be for another moment. I, for one, find the prospect of being moved to and fro by my strongest desire at each point in time quite distressing. I suspect some will find this prospect less disturbing, but I think we would all have a decent chance of finding ourselves in our own version of *The Scorpion and the Frog* fable, resulting in varying degrees of calamity.[20]

Continuing with our thought experiment, now imagine yourself with a slightly different version of the capacity for self-consciousness that you have now. This self-consciousness will allow you to observe all of your mental activity, but no more. You would not be able to intervene in any way. It would be like taking a mental ride on an animal that lacked self-consciousness. This

---

[20] *The Scorpion and the Frog* fable reads as follows:

> A scorpion and a frog meet on the bank of a stream and the scorpion asks the frog to carry him across on its back. The frog asks, "How do I know you won't sting me?" The scorpion says, "Because if I do, I will die, too." The frog is satisfied, and they set out, but in midstream, the scorpion stings the frog. The frog feels the onset of paralysis and starts to sink, knowing they both will drown, but has just enough time to gasp, "Why?" Replies the scorpion: "It's my nature . . ."

(www.aesopfables.com/cgi/aesop1.cgi?4&TheScorpionandtheFrog)

would provide you with a nice seat for watching yourself sting the frog (or be stung by the scorpion), but you could not affect the outcome at any point. Would you want *this* version of self-consciousness? Although it may be a fun and novel experience for a while, it would soon become wholly frustrating, giving new appreciation for the analogy of our bodies being cages.[21] While I think some serious reflection on this version of self-consciousness shows it to be worse than no self-consciousness at all, all that really need be true is that this version is not preferable to the version we actually enjoy. It is the ability to *intervene*, and not just *observe*, that is so central to our understanding of ourselves and our mental lives.

Frankfurt (2006: 4) claims that this capacity is more fundamental to our humanity than our capacity for reason, albeit more inconspicuous, and goes on to describe it as follows:

> It is our peculiar knack of separating from the immediate content and flow of our own consciousness and introducing a sort of division within our minds. This elementary maneuver establishes an inward-directed, monitoring oversight. It puts in place an elementary reflexive structure, which enables us to focus our attention directly upon ourselves.
>
> When we divide our consciousness in this way, we *objectify* to ourselves the ingredient items of our ongoing mental life. It is this self-objectification that is particularly distinctive of human mentality. We are unique (probably) in being able simultaneously to be engaged in whatever is going on in our conscious minds, to detach ourselves from it, and to observe it—as it were—from a distance. We are then in a position to form reflexive or higher-

[21] This point is made nicely in Dalton Trumbo's book *Johnny Got His Gun* about a WWI soldier who, after being wounded by an exploding artillery shell, awakens to find that although he has lost his arms, legs, eyes, ears, teeth, and tongue, his mind functions perfectly (or as perfectly as one's mind could given this situation).

> order responses to it. For instance, we may approve of what we notice ourselves feeling, or we may disapprove; we may want to remain the sort of person we observe ourselves to be, or we may want to be different. Our division of ourselves situates us to come up with a variety of supervisory desires, intentions, and interventions that pertain to the several constituents and aspects of our conscious life.

It is through the lens of our reflective capacity that the presence of first-order desires in persons comes into focus. After noting that some philosophers claim that a person necessarily has a reason to satisfy any desire he has, Frankfurt (2006: 11) states that

> the mere fact that a person has a desire does not give him a reason. What it gives him is a problem. He has the problem of whether to identify with the desire and thus validate it as eligible for satisfaction, or whether to dissociate himself from it, treat it as categorically unacceptable, and try to suppress it or rid himself of it entirely.

It is through this process that we participate in our agency. Thus a person, but not the scorpion, can reflect upon his desire to kill the frog midstream and decide what to do about the problem (i.e., desire) he now finds himself confronted with. All of this plays out in the structure of our wills, which is where personhood is located, through the formation of second-order volitions—the essential feature of persons. Accordingly, being a person entails having an evaluative attitude toward oneself. Persons endorse or repudiate their motives and organize their preferences and priorities.

This participation in our agency is also what constitutes our being active. Recall the passage from Frankfurt (1988: ix) in which he draws our attention to the importance of the active-passive distinction: “The difference between

passivity and activity is at the heart of the fact that we exist as selves and agents and not merely as locales in which certain events happen to occur." Frankfurt notes that we are passive with respect to numerous events within our bodies: the dilation of my pupils, the beating of my heart—these are not actions that *I* perform. The same can be said of our intellectual processes. Thus we can be active, as with "turning one's mind in a certain direction, or deliberating systematically about a problem," or we can be passive, as with "obsessional thought, whose provenances may be obscure and of which we cannot rid ourselves; thoughts that strike us unexpectedly out of the blue; and thoughts that run willy-nilly through our heads" (Frankfurt 1988: 59). We do not actively participate in the occurrence of these latter thoughts; we merely find them occurring within us. The key, in the case of both bodily movements and intellectual processes, is whether these things occur "*under the person's guidance*" (Frankfurt 1988: 72). The person's guidance is achieved through the will and the desires that comprise the will:

> Now a person is active with respect to his own desires when he identifies himself with them, and he is active with respect to what he does when what he does is the outcome of his identification of himself with the desire that moves him in doing it. Without such identification the person is a passive bystander to his desires and to what he does . . . . (Frankfurt 1988: 54)

Specifically concerning the second-order volitions that are essential to personhood, it is impossible for a person to be passive with respect to them, as "they *constitute* his activity" (Frankfurt 1988: 54). Frankfurt (1999: 79) sums up these ideas as follows:

> *[T]he will is absolutely and perfectly active*. In other words, there can be no such thing as a passive willing. All of the movements of my will – for instance, my choices and decisions – are *movements that I make*. None is a mere impersonal occurrence, in which my will *moves without my moving it.* None of my choices or decisions merely happens. Its occurrence *is* my activity, and I can no more be a passive bystander with respect to my own choices and decisions than I can be passive with respect to any of my own actions. It is possible for me to be passive when my arm rises, but I cannot be passive when I raise it. Now every willing is necessarily an action; unlike the movements of an arm, it is only as actions that volitions can occur. Thus, activity is of the essence of the will. Volition precludes passivity by its very nature.

VII. Identification & Wholeheartedness

In setting out Frankfurt's analysis of personhood, I mentioned a person's "identifying" with one of his first-order desires. The act of identifying with a first-order desire was claimed to be the key to making a desire the person's *own* desire in a nontrivial way and, accordingly, essential to being a person at all. The fact, then, that this is a central concept in Frankfurt's account of personhood demands that it be explained.

The act of identifying with a desire is rooted in the phenomenology of human mentality and arises as a consequence of the reflexive nature of our minds. Our reflexive capacity is what allows us to make decisions at all, and identification is a kind of decision. "To make a decision is to make up one's mind. This is an inherently reflexive act, which the mind performs upon itself. Subhuman animals cannot perform it because they cannot divide their consciousness. Because they cannot take themselves apart, they cannot put their minds back together" (Frankfurt 2006: 13). The decision at issue here concerns our psychic raw material—the desires we find within ourselves that

are provided by nature and circumstance. The decision to identify with certain desires and not others depends upon what the person wants himself to be. When a person identifies with a desire, he incorporates it into himself and makes it his own.

> This willing acceptance of [desires] transforms their status. They are no longer merely items that happen to appear in a certain psychic history. We have taken responsibility for them as authentic *expressions of ourselves.* We do not regard them as disconnected from us, or as alien intruders by which we are helplessly beset. The fact that we have adopted and sanctioned them makes them intentional and legitimate. Their force is now our force. When they move us, we are therefore not *passive*. We are active, because we are being moved just by ourselves. (Frankfurt 2006: 8)

It is just this active process of sorting through the desires we find within ourselves that is at the center of volitional capacity. When we identify with some desires and not others, we *internalize* those desires and externalize the others. To internalize a desire is to take it as an authentic expression of ourselves and to give it a place in the ordering of our preferences and priorities. Externalizing a desire, then, will be just the opposite. Although we cannot help having desires that are antithetical to our conceptions of ourselves, we can resolve (i.e., will) ourselves not to let these desires impact our conduct. When we are beset by such desires, we resist them by trying to repress or inhibit them, to dissociate ourselves from them.

> This means that we deny them any entitlement to supply us with motives or with reasons. They are outlawed and disenfranchised. We refuse to recognize them as grounds for deciding what to think or what to do. . . . The fact that we continue to be powerfully moved by them gives them no rational claim. Even if an externalized desire turns out to be irresistible, its

> dominion is merely that of a tyrant. It has, for us, no legitimate authority. (Frankfurt 2006: 10)

An example may help to clarify this point. Suppose I have a friend who is about to achieve a difficult and worthwhile goal. While I am happy for my friend, I also experience some unwelcome jealousy. After taking stock of my desires upon receiving his news, I find three relevant desires: the desire to buy my friend a nice, congratulatory gift; the desire to take my friend out for a nice, congratulatory dinner; and the desire to kill my friend. After a very brief reflection on these desires, I instantly externalize the desire to kill my friend. I am horrified by this desire, try to inhibit it, want it to play no role in either my motivations or actions, and give it no place in the ordering of my projects and priorities. Conversely, the other two desires are ones I decide to internalize by giving them a place in that ordering. After concluding that attempting to satisfy both desires might be a bit much given the occasion and the level of our relationship, I decide that the dinner would be more appropriate and give it a higher priority. However, this does not mean that the desire to give a gift is externalized or rejected. If it turns out that I am unable to coordinate a dinner with my friend, then I should attempt to satisfy the desire to give the gift, as it is a desire within the ordering of my priorities (i.e., a desire that I willingly adopted and sanctioned).

At this point, it is hopefully clear that the concept of identification involves second-order desires that result from the reflexive capacity of our

minds.[22] But what determines whether we identify with a second-order desire in a way that is sufficient to turn it into a second-order volition, which is the essential element of personhood? This hierarchical account of the self seems to leave itself open to the objection that the second-order desire in question is just another desire with no special authority. Why claim that it is constitutive of what the person really wants? As we shall soon see, the search for an even higher-order desire to confirm the authority of this second-order desire will only lead to a fruitless, infinite regress. Moreover, a second-order desire does not become a second-order volition merely because it endorses a first-order desire. In other words, simply *having* a second-order desire that endorses a first-order desire is not enough.

> The endorsing higher-order desire must be, in addition, a desire with which the person is *satisfied*. . . . Identification is constituted neatly by an endorsing higher-order desire with which the person is satisfied. It is possible, of course, for someone to be satisfied with his first-order desires without in any way considering whether to endorse them. In that case, he is identified with those first-order desires. But insofar as his desires are utterly unreflective, he is to that extent not genuinely a person at all. He is merely a wanton. (Frankfurt 1999: 105-06)

It would, of course, be a welcome outcome if we always had higher-order desires with which we were satisfied. Unfortunately, this is not the case. The reflexive capacity of our minds virtually assures this outcome. The ability to reflect upon the contents of our minds makes it statistically unlikely that everyone would always be satisfied with all the elements he finds there. This

---

[22] There is, of course, the possibility that desires of an even higher order should come into play. I will leave this bit of complexity aside here, but will address it below.

fact gives rise to the common occurrence of our wanting to be different than we are. Sometimes this inner conflict even goes beyond the level of competing first-order desires, as in the case of Dexter and Dahmer. For just as there can be conflicts between first-order desires, so too can there be conflicts between higher-order desires.

This possibility is one that must be addressed by any hierarchical account of the self, particularly one that defines personhood as Frankfurt does. Recall that a person must have at least one second-order volition. However, second-order volitions are possible only if the conflict, if any, between his second-order desires is resolved in a manner sufficient to produce a preference concerning which of his first-order desires is to be his will. To the extent that this conflict is unresolved,

> if it is so severe that it prevents him from identifying himself in a sufficiently decisive way with *any* of his conflicting first-order desires, [it] destroys him as a person. For it either tends to paralyze his will and to keep him from acting at all, or it tends to remove him from his will so that his will operates without his participation. In both cases he becomes . . . a helpless bystander to the forces that move him. (Frankfurt 1988: 21)

Conflict of this kind that is wholly within a person's volitional complex is just what it is to be *ambivalent*. There are degrees of ambivalence, but what concerns us here is an ambivalence to such a degree that the person cannot act decisively or finds that fulfilling either of her conflicting desires is substantially unsatisfying (Frankfurt 1999: 99). This level of ambivalence is the result of conflicting volitional movements that are wholly internal to a person's will *and* that are inherently opposed such that they cannot all be

satisfied (Frankfurt 1999: 99). This condition could be due either to being drawn both to and away from the same state of affairs, or to a conflict that makes it impossible for all of the desired states of affairs to be brought about (Frankfurt 1988: 165). The result is the "possibility that there is no unequivocal answer to the question of what the person really wants, even though his desires do form a complex and extensive hierarchical structure" (Frankfurt 1988: 165). His will is unformed due to being inclined in opposite directions such that "it is true of him neither that he prefers one of his alternatives, nor that he prefers the other, nor that he likes them equally" (Frankfurt 1999: 100).

While this description of ambivalence does not make it sound good, it is worth being very clear about why ambivalence is so problematic in terms of personal welfare. If ambivalence is a volitional malady, our wills are comprised of desires, and the correct theory of personal welfare is some form of desire satisfactionism, then ambivalence is a potentially troublesome mental state. Frankfurt (1999: 99) sums up the problem in this way:

> A person is ambivalent, then, only if he is indecisive concerning whether to be for or against a certain psychic position. Now this kind of indecisiveness is as irrational, in its way, as holding contradictory beliefs. The disunity of an ambivalent person's will prevents him from effectively pursuing and satisfactorily attaining his goals. Like conflict within reason, volitional conflict leads to self-betrayal and self-defeat. The trouble is in each case the same: a sort of incoherent greed – trying to have things both ways – which naturally makes it impossible to get anywhere. The flow of volitional or of intellectual activity is interrupted and reversed; movement in any direction is truncated and turned back. However a person starts out to decide or to think, he finds that he is getting in his own way.

Although ambivalence is a problem, it is sometimes warranted. Many a case could be conceived in which being unsure whether to endorse or repudiate a desire would be helpful or wise. Yet this ambivalence will only be good for what it leads to, not in itself. In other words, ambivalence *may* ultimately help produce a better decision, but it could never be a mental state that one could desire for its own sake (Frankfurt 1999: 102). This is easy to see if one imagines an all-encompassing ambivalence such that the person involved is, with regard to every action and motive, always completely unsure of what to do and why to do it. Such an outcome would seem to be an almost perfect state of mental torment that could be equivalent to the hearing-voices-in-my-head scenario that is often portrayed as being the paradigmatic case of mental torment. Given this description of ambivalence, then, it is true of persons that they are not ambivalent about ambivalence. We fully and unequivocally desire to be volitionally unified (i.e., desire to avoid the volitional division that is ambivalence) (Frankfurt 1999: 106).

Being volitionally unified is what Frankfurt (2004: 96) calls being *wholehearted*. When a *person* is not ambivalent, then the person is wholehearted. However, the conflicting desires need not change in either their objects or their intensity. What being wholehearted requires is that the *person* become finally and unequivocally clear as to what side of the conflict he is on (Frankfurt 2004: 91). When this happens wholly within the volitional complex, the opposed desire is externalized and is now opposed by the *person* and not by just another competing desire (Frankfurt 2004: 91). In this way, being

wholehearted is compatible with virulent psychic conflict as long as the person is uninhibitedly and unqualifiedly on one side of the conflict or the other. Wholeheartedness is a kind of self-satisfaction whereby a person willingly accepts and endorses his own volitional identity (Frankfurt 2004: 96). There is no equivocation or resistance from the parts of himself with which he identifies, and this lack of division within the will means that the will the person has is the will he wants.

And although I think the term "wholehearted" is a good one for the concept in question, there is some concern about confusion given its popular usage. Wholeheartedness is a structural characteristic and "not a measure of the firmness of a person's volitional state, or of his enthusiasm. What is at issue is the organization of the will, not its temperature" (Frankfurt 1999: 100). Wholeheartedness is just the basis for practical rationality, which renders our practical lives coherent. This means that what counts here is the *quality* of our wills, not the quantity of its objects. Being wholehearted with respect to one desire is also consistent with assigning a higher priority to another desire that happens to contingently conflict with it, as in the case of the successful friend discussed above (Frankfurt 1999: 103). It should also be noted that a person's will can be defeated without disrupting its unity, as in the case of Dexter's identifying himself with the second-order volition not to kill, yet being overcome by the externalized effective first-order desire to kill. Finally, being wholehearted does not entail being closed-minded. "The wholehearted person need not be a fanatic. Someone who knows without qualification where he

stands may nonetheless be quite ready to give serious attention to reasons for changing that stand. There is a difference between being confident and being stubborn or obtuse" (Frankfurt 2004: 95 n.7).

One final aspect of wholeheartedness merits an extended discussion in order to avoid confusion. Most of the ways wholeheartedness has been described above are just the effects of being wholehearted (e.g., an undivided will), but we should be very clear about what *produces* wholeheartedness. Being wholehearted consists in being fully satisfied that the desires in question, rather than others that inherently conflict with them, "should be among the causes and considerations that determine [one's] cognitive, affective, attitudinal, and behavioral processes" (Frankfurt 1999: 103). Satisfaction entails an absence of restlessness and resistance such that the person has no active interest in bringing about any change. However, this does not mean that the person could not be satisfied with any change in his condition. While it is almost certainly true that a person would be satisfied with an improved condition, this possibility does not engage his concern because being satisfied does not require an all-encompassing drive to maximize (Frankfurt 1999: 103). In fact, a person may even be satisfied with a condition *inferior* to the one in which he now finds himself. The bottom line is that being satisfied is *having no interest* in making changes because psychic elements of certain kinds *do not occur* (Frankfurt 1999: 105).

It is important to draw attention to this fact because it explains why there is no danger of a problematic regress in our hierarchical structure of the

will. If being satisfied with a desire required some other intentional act in order to make it the case that the person was satisfied, then a regress problem would ensue:

> Suppose that being satisfied did require a person to have, as an essential constitutive condition of his satisfaction, some deliberate psychic element – some deliberate attitude or belief or feeling or intention. This element could not be one with which the person is at all dissatisfied. How could someone be wholehearted with respect to one psychic element by virtue of being halfhearted with respect to another? So if being satisfied required some element as a constituent, satisfaction with respect to one matter would depend upon satisfaction with respect to another; satisfaction with respect to the second would depend upon satisfaction with respect to still a third; and so on, endlessly. Satisfaction with one's self requires, then, no adoption of any cognitive, attitudinal, affective, or intentional stance. It does not require the performance of a particular act; and it also does not require any deliberate abstention. Satisfaction is a state of the entire psychic system – a state constituted just by the absence of any tendency or inclination to alter its condition. (Frankfurt 1999: 104)

Finally, satisfaction must be uncontrived and reflective. Any attempt to simulate satisfaction through an intentional effort would bring us right back to the regress issue. The satisfaction at issue must be integral to the person's psychic condition and not due to any effort by the person to make it so (Frankfurt 1999: 104). However, the absence of the psychic elements in question must nonetheless be reflective.

> In other words, the fact that the person is not moved to change things must derive from his understanding and evaluation of how things are with him. Thus, the essential non-occurrence is neither deliberately contrived nor wantonly unselfconscious. It develops and prevails as an unmanaged consequence of the person's appreciation of his psychic condition. (Frankfurt 1999: 105)[23]

[23] Frankfurt (1999: 105 n.16) continues to clarify this concept as follows:

The value of being wholehearted results from the fact that it fixes a serious problem for us. Saint Augustine (*Confessions*: 8.9) noted that while it is "no strange phenomenon partly to will to do something and partly to will not to do it," he claimed that this was a "disease of the mind" perhaps inflicted on us by god as a punishment for original sin. Spinoza (1981: 199) made a very similar claim about the deleterious effect of a divided will when he claimed that the sort of satisfaction or self-approval described above "is in reality the highest object for which we can hope." According to Frankfurt (2006: 18), this *is* the highest thing for which we can hope:

> Perhaps because it resolves the deepest problem. In our transition beyond naive animality, we separate from ourselves and disrupt our original unreflective spontaneity. This puts us at risk to varieties of inner fragmentation, dissonance, and disorder. Accepting ourselves reestablishes the wholeness that was undermined by our elementary constitutive maneuvers of division and distancing. When we are acquiescent to ourselves, or willing freely, there is no conflict within the structure of our motivations and desires. We have successfully negotiated our distinctively human complexity. The unity of our self has been restored.

> Being or becoming satisfied is like being or becoming relaxed. Suppose that someone sees his troubles recede and consequently relaxes. No doubt it is by various feelings, beliefs, and attitudes that he is led to relax. But the occurrence of these psychic elements does not constitute being relaxed, nor are they necessary for relaxation. What is essential is only that the person stop worrying and feeling tense.

VIII. CARING

The concept of caring has played a prominent role in this project thus far. I have argued for both the Principle Concerning Caring (PCC)[24] and the Internalist Principle' (IP').[25] As both of these principles rely heavily on the concept of caring, it is time to clarify this concept and demonstrate it adequately handles the reliance placed upon it.

Before turning to an examination of the concept of caring, I would like to adopt a terminological stipulation that should both help our understanding of caring and help to avoid some potentially awkward phrasing. Although the phrases "he cares about X" and "he regards X as important to himself" are, perhaps, not perfectly synonymous, I will treat them as substantially equivalent from this point forward. This should not create any confusion, as I think it is fair to say both that people care about things they regard as important to themselves and that regarding something as important also entails caring about that thing. As Frankfurt adopts this same stipulation, it will have the added benefit of helping us to understand his arguments (Frankfurt 1999: 155-56).

So what does it mean, exactly, when a person claims to *care* about something? It appears as though desire must be at least part of the right

---

[24] The Principle Concerning Caring states: If X does not care about anything and could not be made to care about anything in the future, then it is not possible for X's personal welfare to be affected going forward.

[25] The Internalist Principle' states: The value of a life (or part of a life) for the one who lives it is determined to a significant degree by what the person in question cares about.

account. This is because it does not seem coherent to claim both that a person cares about X and that this person has no desires pertaining to X. Accordingly, it is a necessary condition of caring about an object to have at least one desire concerning that object.

It is possible, then, that this is both a necessary and sufficient condition for caring. In other words, when we say we care about something, are we merely claiming that we have some desire concerning it? No. The reason is that it looks to be quite coherent to claim that we want X, but that we do not really care about X or our desire for X. For example, I may see an advertisement for a Slurpee and then find that I now have a slight desire for one. However, after a bit of reflection, I conclude that this desire is not one that I care about. And it need not be because I have some competing desire that I care about more, like a desire to avoid junk food or to avoid impulse buying. I could simply and coherently decide that it is true both that I have this new desire for a Slurpee and that this desire is not one that I care about satisfying.

While it may be true that a person does not care about a fairly weak desire, perhaps it is true that a person must necessarily care about a desire of a certain intensity level. In other words, maybe caring is simply having a sufficiently strong desire.[26] The most obvious, and perhaps only, argument

---

[26] The argument that we saw earlier from William James (1948: 73) might produce this conclusion: "Take any demand, however slight, which any creature, however weak, may make. Ought it not, for its own sole sake, to be satisfied? . . . Any desire is imperative to the extent of its amount; it *makes* itself valid by the fact that it exists at all."

that would support this conclusion would go as follows.[27] An unsatisfied desire involves, by definition, frustration. Frustration is unpleasant. There is a presumption in favor of minimizing unpleasantness. Therefore, there is a presumption in favor of satisfying desires. One objection to this argument would be to deny that all desire frustrations involve unpleasantness. Whatever the merits of this claim, it is not applicable here, as we are only dealing with the class of relatively intense desires. Accordingly, we could stipulate that frustrations of desires in this class do involve some degree of unpleasantness. Does this mean, then, that caring consists in having desires of a sufficient level of intensity?

No. One way to see this is to understand that caring is an evaluative attitude and that differences in the intensities of our desires may be due to all sorts of things that are independent of this evaluative attitude. Indeed, if this evaluative attitude simply measured the relative intensity of desires, it would not be much of an "evaluative attitude" at all. In that case, it would be much more appropriate to call caring a score-keeping function. The bottom line is that what a person thinks of a desire need not be strictly determined by the intensity of the desire; it is just another factor to consider. Another way to see this is to look back at the argument based on James's claim. The only conclusion actually warranted by the premises is that there is a presumption in favor of satisfied desires over frustrated ones. This, however, does not exhaust the logical space when it comes to desires. Even if we suppose that

[27] This argument would work only for desire satisfactions and frustrations that the desirer is aware of.

the frustration of a desire entails some unpleasantness, this unpleasantness can be avoided without satisfying the desire. This is because a person may *give up* or, in some other way, *lose* a desire. This is, in fact, the preferred course of action for some people when it comes to desires of any intensity. Therefore, James's argument fails to establish that we must care about intense desires because we must care about avoiding the unpleasantness of frustration.

The inference that we must care about *satisfying* an intense desire is fatally flawed in yet another way. Not only may we not care about satisfying an intense desire, we may actually care about *frustrating* an intense desire. For example, a drug addict has, by definition, an intense desire for his drug of choice. However, he may struggle against his addiction and, thus, have a strong higher-order desire for his first-order desire for the drug to be frustrated. This last point helps clarify another way in which we can see that caring does not reduce to desires. Suppose I am watching television in 1976 and that I prefer watching *Wide World of Sports* to watching the only other alternatives of *American Bandstand*, *Soul Train*, or *The Electric Company*. Can we infer that I care about watching *Wide World of Sports* to any of the other three alternatives? No. As noted above, the fact that I prefer one to the other does not entail that I care about it more, because it does not entail that I care about it at all (Frankfurt 1999: 157).

Well, if the intensity of desires does not determine what it is we care about, perhaps caring is desiring something *and* believing the thing sought is

intrinsically valuable such that it is pursued as a final end. Of course, intrinsic value alone will not suffice since we have already determined that desire is at least part of the correct account of caring:

> Even if a person believes that something has considerable intrinsic value, he may not regard it as important to himself. In attributing intrinsic value to something, we do perhaps imply that it would make sense for someone to desire it for its own sake—that is, as a final end, rather than merely as a means to something else. (Frankfurt 2004: 12)

Further:

> Despite his recognition of its value, it may just not appeal to him; and even if it does not appeal to him, he may have good reason for neither wanting it nor pursuing it. Each of us can surely identify a considerable number of things that we think would be worth doing or worth having for their own sakes, but to which we ourselves are not especially drawn and at which we quite reasonably prefer not to aim. (Frankfurt 1999: 158)

Moreover, even if we do attribute intrinsic value to some object, *and* we desire it for its own sake and pursue it as a final end, it still cannot be presumed that we care about it. This is because intrinsic value is merely a *type* of value and, thus, does not reflect the *amount* of value we attribute to the object (Frankfurt 2004: 13). We may, in fact, attribute a small amount of intrinsic value to something, as is the case with many of the inconsequential pleasures we seek. I may want doughnuts, for example, and want the doughnuts simply for the intrinsic value of the pleasure they bring me. This does not mean that I care about the doughnuts or about my desire for them. "There is no incoherence in appraising something as intrinsically valuable, and pursuing it actively as a

final end that is worth having in itself, and yet not caring about it" (Frankfurt 1999: 159).

If caring involves desires of a certain kind, and if strong desires and believing something to be intrinsically valuable are not enough, what, then, does it mean to care about something? Perhaps the best way to understand caring is first to examine what it means *not* to care about something. As we have seen, it is possible to have a desire for X and not care about either X or the desire for X. The most reasonable explanation for this is that there is a lack of commitment to this desire such that the person would be quite willing to give it up.

> In an idle moment, we may have an idle inclination to flick away a crumb; and we may be quite willing to be moved by that desire. Nonetheless, we recognize that flicking the crumb would be an altogether inconsequential act. We want to perform it, but performing it is of no importance to us. We really don't care about it at all.
>
> What this means is not that we assign it a very low priority. To regard it as truly of no importance to us is to be willing to give up having any interest in it whatever. We have no desire, in other words, to continue wanting to flick away the crumb. It would be all the same to us if we completely ceased wanting to do that. When we do care about something, we go beyond wanting it. We want to *go on* wanting it, at least until the goal has been reached. Thus, we feel it as a lapse on our part if we neglect the desire, and we are disposed to take steps to refresh the desire if it should tend to fade. The caring entails, in other words, a commitment to the desire. (Frankfurt 2006: 18-19)

Therefore, caring requires more than just having a desire and more than accepting, approving of, or endorsing a desire. Caring requires wanting the desire *sustained* (Frankfurt 2004: 16). Wanting a desire sustained should be thought of as being disposed to be active in ensuring that the desire is not

abandoned or neglected (Frankfurt 1999: 162). This focus and attention on the desire stem from the fact that this desire is one with which the person *identifies* himself, and which he accepts as expressing what he really wants (Frankfurt 2004: 16).

At this point, it should be clear that caring is a matter of *will*. This is because we are dealing with a subset of higher-order volitions (i.e., a second-order desire such that the person wants the first-order desire in question to be the desire that moves him effectively to act). Specifically, caring consists in having a higher-order volition that the person also wants sustained (Frankfurt 2004: 16). An objection to locating caring within the will might claim that caring is more properly a function of a person's beliefs or feelings. While these are certainly germane to the discussion of what one cares about, Frankfurt (1999: 110-11) argues that there is more to the story:

> [Caring] is not primarily either a cognitive or an affective matter. Cognitive and affective considerations are its sources and grounds. But though it is based on what a person believes and feels, caring is not the same as believing or feeling. Caring is essentially volitional; that is, it concerns one's will. The fact that a person cares about something or considers it important to himself does not consist in his holding certain opinions about it; nor does it consist in his having certain feelings or desires. His caring about it consists, rather, in the fact that he *guides* himself by reference to it. This entails that he purposefully direct his attention, attitudes, and behavior in response to circumstances germane to the fortunes of the object about which he cares. A person who cares about something is, as it were, invested in it. By caring about it, he makes himself susceptible to benefits and vulnerable to losses depending upon whether what he cares about flourishes or is diminished. We may say that in this sense he *identifies* himself with what he cares about.

Desires that a person does not care about, on the other hand, may still persist. After all, one cannot simply be rid of first-order desires that one does not want through an exercise of the will or otherwise. However, these desires are denied any place at all in the *person's* order of preferences—not simply assigned a lower priority—in an effort to cease to be moved by their appeal.[28] By alienating a desire in this way, the person is aiming at disenfranchising the desire, so to speak (Frankfurt 1999: 161). It is only when a desire persists through *the person's own doing* that it can be properly claimed that the person cares about it, although this volitional activity need not be fully conscious or explicitly deliberate (Frankfurt 1999: 160).

Having established that to care about something means being motivated by a concern for it in the way just described, we should examine the objects of our caring more closely. Here, it seems relatively clear, the object of our caring can be almost anything: "a life, a quality of experience, a person, a group, a moral ideal, a nonmoral ideal, a tradition, whatever" (Frankfurt 2006: 40). It should also be noted here that caring about something need not be limited to the typical nurturing connotation associated with this term. The concern necessary for caring "may be positive or negative: hatred or love, a desire to possess or a desire to avoid, an interest in sustaining the object or in destroying it" (Frankfurt 1999: 93).

---

[28] In other words, the *person* does not identify with these desires, even though the desire may still persist and, obviously, be located within the person's body.

While these preliminary remarks concerning the objects of our caring are all well and good, they do not get us any closer to the 800-pound gorilla in the room. This is, of course, the normative question concerning what we *should* care about. For everyone trying to avoid the unexamined life that Socrates warned us about,[29] this is a question that inevitably arises. So, what should we care about? This question is normally interpreted as inquiring about the identity of things that are inherently and objectively important such that they are worth caring about. Is there a sufficient basis for establishing something as being genuinely important in itself, regardless of what anyone thinks about it? No. "There can be no rationally warranted criteria for establishing anything as inherently important" (Frankfurt 2006: 22).[30] There are at least two ways to see this. First, an examination of *objects* that might qualify will demonstrate this fact—a bottom-up approach to the question. Second, a closer examination of the *question* will also establish this fact—a top-down approach.

The bottom-up approach to the question of inherent importance begins with a survey of objects that might qualify for this distinction and then would try to show how they were, in fact, inherently important. The most obvious way to do this, and the only reasonable way, is to ground the justification for

---

[29] Plato (*Apology*: 38a).

[30] Two points bear mentioning here. First, this claim need not be true for my overall project to succeed. The reason to include it, then, relates to the second point. And that point is, to the extent this claim is true, it helps to further weaken the case for objective-list theories, which in turn may increase the plausibility for other theories of welfare (including the one put forward in Chapter Five).

our believing that certain things are inherently important in judgments about the value of those objects. Let us start with a particular value judgment and then broaden our scope to examine value judgments in general. If you were to ask a sufficiently large group of people what you *should* care about, I suspect the most common answer would be either other persons in general or some subset thereof. If you should care about other people, then this seems to entail that you should care about morality since morality deals with how one should conduct oneself in affairs that affect other people.[31] Now the strictures of morality are only what they are in virtue of a value judgment; namely, that other persons are of some value.[32] So morality tells us that other people do have value, but does this make other people necessarily important to us or, perhaps, *the* most important thing? No.

> Now why should *that* be, always and in all circumstances, the most important thing in our lives? No doubt it is important; but, so far as I am aware, there is no convincing argument that it must invariably override everything else. Even if it were entirely clear what the moral law commands, it would remain an open question how important it is for us to obey those commands. We would still have to decide how much to care about morality. Morality itself cannot satisfy us about that (Frankfurt 2006: 28).

The bottom line here is that a person who does not value other people is not making any sort of logical mistake. It may be unthinkable and abhorrent to

---

[31] Limiting the scope of morality to persons here strengthens the objection that I am arguing against. However, expanding the scope to all sentient beings, the proper ambit in my opinion, would do nothing to weaken the argument.

[32] This can easily be seen if one tries to imagine the content of morality if other persons have no value whatsoever. This point has been made by others in their discussions of ethical egoism. They question whether ethical egoism is a moral theory at all, as opposed to just a system of practical reasoning, because it fails to assign any value to the interests of others.

us, but there are no contradictions to be found here. This is just another in the long line of unfortunate facts about the reality we find ourselves in, but of course the fact that we may want something to be different than it is does not make it so or, for that matter, even more likely that it is so.

What, then, about value judgments in general? As was touched on earlier in the chapter, these suffer the same fate as the value judgments of morality, and for the same reason. Using a bit of imagination, you should be able to come up with a relatively long list of objects, activities, and states of affairs that you consider to be inherently valuable, or worthy of being pursued for their own sakes, but that are not important to you and about which you do not care. Other things, perhaps even things that we freely recognize have less value than some of the things on the list of valuable things that you do not care about, are more important to us.[33] What we should care about simply cannot be based on judgments about value. Simply stated, "From the fact that we consider something to be valuable, it does not follow that we need to be concerned with it" (Frankfurt 2006: 27). Accordingly, importance is never inherent, with one exception that need not detain us here.[34]

---

[33] A prime example of this might be a football coach who, although he has dedicated his life to football, might freely acknowledge many other things as being more valuable. Of course, this is not meant to unfairly single out sports. One could substitute all sorts of other career paths, artistic endeavors, etc., to make the same point.

[34] The exception Frankfurt makes here need not detain us because it does not undermine my argument. If it has any effect at all, this exception strengthens my earlier argument concerning the definition of personhood:

> What is important to someone depends upon what he cares about. This might well seem to imply that there can be nothing

Perhaps one might not be persuaded by this bottom-up approach to the question of what one should care about. The top-down approach will produce the same conclusion, albeit from the opposite direction. The problem with the

---

> whose importance to anyone is inherent in it. The fact that something is important to a person is invariably a function of that person's feelings, attitudes, and intentions. Considered just in itself, entirely apart from any consideration of what the person in question cares about, nothing can be said either to be or not to be important to him. For the extent to which something is important to him depends essentially upon considerations other than its own inherent characteristics alone.
>
> However, there is an exception to this principle that the importance of anything depends upon considerations outside itself. People are capable of making themselves important. If a person is important to himself, then this importance is manifestly self-endowed; for he is important to himself simply by virtue of the fact that he cares about himself. In this one case, the source of importance lies in the characteristics of the object. The importance of a person to himself is unique in that it is in no way extrinsic. From this it follows, of course, that someone who enjoys this importance cannot be deprived of it by anything other than himself. Only if a person does not care about himself can he fail to be important to himself.
>
> Now whether a person is or is not important to himself may appear to be a straightforwardly contingent matter. It depends just upon whether he cares about himself; and surely, it seems, he might either do so or not. The importance of a person to himself is clearly intrinsic, in that it depends exclusively upon his own characteristics. If the characteristics upon which it depends are indeed contingent, however, then it is also conditional. But is it actually possible for there to be a person who does not care about himself? Perhaps caring about oneself is essential to being a person. Can something to whom its condition and activities do not matter in the slightest properly be regarded as a person at all? Perhaps nothing that is entirely indifferent to itself is really a person, regardless of how intelligent or emotional or in other respects similar to persons it may be.
>
> Suppose that the sort of reflexivity in question here were, indeed, a conceptually essential characteristic of persons. Then there could not possibly be a person of no importance to himself. To be sure, the importance of a person to himself would still be conditional. But no person could fail to meet the required condition. (Frankfurt 1999: 89-90)

question "What should I care about?" is that it is doomed from the start because it is *systematically inchoate* (Frankfurt 2004: 25). What this means can be discerned by taking a closer look at the question. In particular, let us examine what one would need in order to *answer* the question. Carrying out a rational evaluation of the various things that one should care about would require one to know what evaluative criteria to employ and how to employ them. Specifically, one would need to know what considerations count in favor of caring about something, what considerations count against caring about something, and the relative weights of each (Frankfurt 2004: 24). An evaluation of potential objects of caring without specifying these things cannot even be called an evaluation—at least not a rational one—and, thus, cannot succeed because it cannot even begin. This explains why the question is systemically inchoate. Alternatively, the question could be described as suffering from a vicious sort of circularity, which is necessarily the case if the question is not incomplete. This is because fully formulating the question in a way that would allow it to be answered requires having already settled upon the judgments at which the question aims to provide. In other words, identifying the question is tantamount to answering it since the answer is that one should care about the things that best satisfy the criteria set out in the question itself (Frankfurt 2004: 25).[35]

---

[35] David Sobel (1994: 808) makes a similar point in the context of discussing the difficulties associated with full-information accounts of well-being:

> The narrative unity of a life can provide the context to make sense of choosing one option over another, but this context is

This seems to be a discouraging result. After all, this does seem to be the most basic inquiry that a person trying to lead an examined life could make. Knowing what one should care about, and being able to construct a plan of life on that basis, seems central to that effort. So where does that leave us? It turns out that we still can sensibly ask what we *should* care about, but only after we have an answer to the question of what we actually *do* care about. In other words, the *factual* question of what we do care about must precede the *normative* question of what we should care about (Frankfurt 2006: 23). This conclusion becomes clear if we examine the case of a person who does not care about anything. If a person cares about nothing, then a rational inquiry into what he should care about cannot begin because the fact that he cares about nothing entails that there is nothing that can count with him as a reason for caring about one thing over another or caring about anything at all (Frankfurt 2004: 26). And if he truly does not care about anything, then he will not care about this, either. However, if a person does care about something, then it may be possible for him to discover other things that he should care about as well. For example, suppose a person cares about his health, but does not care about the nutritional value of the food he consumes. If this person inquired

---

> significantly dropped when we are choosing between lives rather than from within them. It is not just the different life paths that we could lead that are to be chosen between, it is also who we are to be; what kind of person we want to be who is having these different lives. Without anything like the context provided by our actual lives, the chooser becomes disorientingly "unencumbered."

The context Sobel speaks of is provided by what we do care about. If we are unaware of what we care about, it is impossible to get a foothold in the question of what we should care about in the possible alternate lives we could lead.

into what he should care about, he may discover that he *should* care about the nutritional value of his food on the basis of a fact that he does care about—namely, his physical health. For unless he does care about his health (or something else that pertains to the ingredients in his food), then there would be nothing that would count for him as a reason that he should care about what is in the food he eats.[36]

Needing to answer the factual question of what we do care about before the normative question of what we should care about is better than not being able to coherently ask the normative question at all, but this result is still unsatisfying. After all, we wanted to know what we should care about, and we were reduced to simply asking what we do care about. The unsatisfying nature of this conclusion stems from the notion that merely knowing how things actually are—in this case what we do care about—seems to do nothing in terms of justifying them or giving us a reason to accept them (Frankfurt 2004: 27-28). Justification and acceptance were the main considerations in asking what we should care about in the first place. However, as this section has shown repeatedly and in a variety of ways, it is the normative question that misses the point, not the factual question. Demonstrating, from the ground up, what we

[36] While it is possible that there could be someone who literally did not care about anything, this is largely a theoretical consideration. Nearly everyone cares about something. In fact, the things people care about overlap to a significant degree, although the ordering of these things does vary a great deal. So people care about their physical and mental health, family, friends, hobbies, livelihood, etc. This is not a coincidence, but rather is due to the fact that human nature and the basic conditions of human life are not subject to a great deal of variation or change (Frankfurt 2004: 27).

should care about—what Frankfurt calls the "pan-rationalist fantasy"—is incoherent and must be abandoned (Frankfurt 2004: 28).[37] Caring does not require proof or definitive arguments, but rather clarity concerning what we do care about. And, as we saw earlier in this chapter, hopefully our clarity is accompanied by our being wholehearted in the pursuit of those things.

From what has been said about the primacy of the factual question concerning what we care about, there are likely many misconceptions lurking. I will try to correct a couple of the most glaring ones here. First, the search for what we actually do care about may seem to imply that we are aiming at a fixed target in that what we care about does not change. This is not the case, as even a cursory examination of persons over time would reveal. What we care about is a function of generic human nature, in most cases, and of a person's own particular makeup and experience. And as our makeup can change and our experience does change, what we care about can, and almost always does, change along with these two things. The good news here is that this outcome meshes well with the overwhelming weight of the empirical evidence. The bad news is that we cannot simply attempt to figure out what we care about and then be done with it once and for all.

Another likely misconception would be that, given that we are only performing a factual inquiry on ourselves, there would be little to no chance

---

[37] This outcome is analogous to a broader problem with philosophy in general. Unless we have some fixed starting points in philosophy, it is impossible even to begin. For example, without at least the minimal principle of noncontradiction, one could not hope to establish a premise, much less a conclusion.

that we could get this wrong. This is not the case. There are two categories of mistakes that people can and do make in this regard. The first involves the objects of our caring themselves and our understanding of them. One way a lack of understanding of an object we care about could cause us to get it wrong is when the object turns out to be different than we initially thought it was. I have personally witnessed this in the legal profession. Some people care about becoming a lawyer only to discover that the practice of law is different than they thought it would be, which causes them to no longer care about being a lawyer. It may also turn out to be the case that, as we learn more about the things we care about, we discover that they conflict with each other in some fundamental way. One might imagine a person who cares about both science and a particular religion. Over time he comes to understand that these two things cannot be reconciled. Perhaps this realization will lead him to the conclusion that he in fact cares more about one than the other so that his internal conflict might be resolved. The bottom line here is that we may not understand the objects we care about such that they turn out to be different than we thought or that the requirements and consequences of caring about them turn out to be different than we thought (Frankfurt 2006: 49).

The second general category of mistakes one can make concerning caring involves not the objects, but ourselves. Just as we can fail to understand the objects we care about, we can fail to understand ourselves. Really understanding ourselves to such a degree to allow us to be clear about what we

care about is no small feat. Frankfurt (2006: 49-50) sums up the difficulty quite well:

> Our motives and our disposition are notoriously uncertain and opaque, and we often get ourselves wrong. It is hard to be sure what we can bring ourselves to do, or how we will behave when the chips are down. The will is a thing as real as any reality outside us. The truth about it does not depend upon what we think it is, or upon what we wish it were.[38]

In terms of the things we can get wrong when it comes to what we care about, this is the end of the list. In fact, it is hard to imagine what else we could get wrong that would not involve something metaphysically far-fetched. *There is simply nothing else to get right* once we understand both the objects we care about and ourselves. As we saw earlier in this section, what we care about cannot be demonstrated by impersonal considerations that all rational agents would accept. This outcome helps to highlight another unfortunate fact of life. Since what we care about cannot be demonstrated in this way, it may, and very often does, turn out that certain conflicts between people involving what they care about are irreducible. Neither party can be shown to have

---

[38] This point helps to illuminate the fact that deciding to care about something is not tantamount to actually caring about it (Frankfurt 1988: 84). By deciding to care about something, a person merely forms an intention to care about it, which may or may not be fulfilled. As Frankfurt (1988: 84) rightly points out, a "decision to care no more entails caring than a decision to give up smoking entails giving it up." The issue here is not primarily that the person will change his mind or that he will forget his decision; rather, it concerns the distinct possibility that he will be *unable* to carry out his decision (Frankfurt 1988: 84). Should this occur, it would just be another manifestation of the person's not knowing himself well enough such that, when the time comes to follow through on his decision, he finds that his decision does not accurately reflect what he cares about and the relative priorities of those things. In short, the problem of what to care about cannot be overcome simply by making a decision.

made a mistake, which leaves little hope for a resolution that both parties can accept (Frankfurt 2006: 50). Once again, the fact that we might prefer a different outcome does not even make it slightly more likely that this is so.

In previous chapters, an intuitive sense of the concept of caring was used extensively to argue against both hedonistic and objective-list theories. In this chapter, the concept of caring has been thoroughly defined and illustrated. Through this process, the general importance of caring to persons should have become clear. However, it is worth being as clear and explicit as possible about the importance of caring before turning to other matters.

First, it is through caring that we infuse the world with importance (Frankfurt 2004: 23). Since caring about an object is what makes that object important to the person who cares about it, it is only through caring that anything *can* be important to a person at all. This fact, however, does nothing to taint the objects we care about. It certainly does not do so for those people who do not know it, but it also does not do so for the people who do know it. This is because caring entails that we regard the objects of our caring as valuable in themselves, despite the fact that our viewing objects in this way cannot be established independently of our caring (Frankfurt 2004: 56).

Here one might object to the idea of importance mattering at all. It seems like a relatively uncontroversial proposition, but is it? Why is it better for a person (assuming the person cares about himself and his welfare) to care about things—thus making those things important—than it is to care about nothing at all? There are two primary reasons why this is the case. The first

involves final ends. In short, caring gives us our final ends by being the originating source of terminal value (Frankfurt 2004: 55). Given the arguments earlier in this chapter, that point should be all but obvious by now. Nonetheless, one might still object to this line of argumentation on the grounds that final ends do not matter any more than importance does. What then?

A person without important final ends would, according to Frankfurt, inevitably be led to boredom, which is a serious problem in its own right. To see why this is so, a closer examination of final ends is in order. Aristotle (*Nicomachean Ethics*: 1094a) claimed that desire is "empty and vain" unless "there is some end of the things we do which we desire for its own sake." In other words, if *everything* we do is done merely for the sake of doing something else, then there seems to be no real point or foundation for any of it. Accordingly, it would be difficult to be genuinely satisfied by any of the things we do since our sequence of actions will always be unfinished, by hypothesis. This, in turn, will cause us to lose interest in what we do (Frankfurt 2004: 52-53).

Frankfurt (1999: 88-89) does an excellent job of describing the problem that results from this condition:

> A life in which it were actually the case that nothing was important would be, by hypothesis, a life without important final ends. It follows that it would be a life without meaningful activity. Anyone who lived that life would be indifferent and unengaged with respect to whatever it might be that he did. Furthermore, he would be bored. I believe that the avoidance of boredom is a very fundamental human urge. It is not a matter merely of distaste for a rather unpleasant state of consciousness. Being bored entails a reduction of attention; our responsiveness to conscious stimuli flattens out and shrinks; distinctions are not noticed and not

> made, so that the conscious field becomes increasingly homogeneous. The general functioning of the mind diminishes. It is of the essence of boredom that it involves an attenuation of psychic liveliness. Its tendency is to approach a complete cessation of significant differentiation within consciousness; and this homogenization is, at the limit, tantamount to the cessation of conscious experience altogether.
>
> A substantial increase in the extent to which we are bored undermines the very continuation of psychic activity. In other words, it threatens the extinction of the active self. What is manifested by our interest in avoiding boredom is therefore not simply a resistance to discomfort but a quite elemental urge for psychic survival. It is natural to construe this as a modification of the more familiar instinct for self-preservation. It is connected to "self-preservation," however, only in an unfamiliarly literal sense – in the sense of sustaining not the *life* of the organism but the *persistence of the self*.

Thus we arrive at the second reason why caring about something is essential to having a good life, and it involves the issue of personal identity. Needless to say, I do not have the space here to set out and defend a theory of personal identity—a topic that would require another dissertation. However, there are clearly issues of personal identity lurking here. And while there are strong claims regarding the relationship between Frankfurt's account of the will and personal identity, I am proposing a fairly modest claim: It is at least plausible that a particular person no longer exists *qua* that person if his will is either removed or completely changed.

Although Frankfurt never fully addresses the topic of personal identity, some of his remarks do lend support to my claim. Consider what Frankfurt (1988: 83-84) says about the "temporal characteristics" of caring:

> The outlook of a person who cares about something is inherently prospective; that is, he necessarily considers himself as having a future. On the other hand, it is possible for a creature to have

> desires and beliefs without taking any account at all of the fact that he may continue to exist.
>
> Desires and beliefs can occur in a life which consists merely of a succession of separate moments, none of which the subject recognizes – either when it occurs or in anticipation or in memory – as an element integrated with others in his own continuing history. When this recognition is entirely absent, there is no continuing subject. The lives of some animals are presumably like that. The moments in the life of a person who cares about something, however, are not merely linked inherently by formal relations of sequentiality. The person necessarily binds them together, and in the nature of the case also construes them as being bound together, in richer ways. This both entails and is entailed by his own continuing concern with what he does with himself and with what goes on in his life.
>
> Considerations of a similar kind indicate that a person can care about something only over some more or less extended period of time. Desires and beliefs have no inherent persistence; nothing in the nature of wanting or of believing requires that a desire or a belief must endure. But the notion of guidance, and hence the notion of caring, implies a certain consistency or steadiness of behavior; and this presupposes some degree of persistence. A person who cared about something just for a single moment would be indistinguishable from someone who was being moved by impulse. He would not in any proper sense be guiding or directing himself at all.

The most important passage from this section for purposes of my personal identity claim above is Frankfurt's claim that *there is no continuing subject* when the appropriate temporal recognition is lacking. This could occur in the case of the boredom described above or, *a fortiori*, in the case where the will is removed altogether.

Next, consider Frankfurt's (1999: 162) remarks concerning a situation wherein we cared about nothing:

> In that case, we would be creatures with no active interest in establishing or sustaining any thematic continuity in our volitional lives. We would not be disposed to make any effort to maintain any of the interests, aims, and ambitions by which we are from time to time moved.

> Of course, we would still be moved to *satisfy* our desires; that is irreducibly part of the nature of desire. We might also still want to have certain desires, and to be motivated by them in what we do; and we might want not to have certain others and want not to be moved by them to act. In other words, our capacity for higher-order desires and higher-order volitions might remain fully intact. Moreover, some of our higher-order desires and volitions might tend to endure and thus to provide a degree of volitional consistency or stability in our lives. From our point of view as agents, however, whatever coherence or unity might happen to come about in this way would be merely fortuitous and inadvertent. It would not be the result of any deliberate or guiding intent on our part. Desires and volitions of various hierarchical orders would come and go; and sometimes they might last for a while. But in the design and contrivance of their succession we ourselves would be playing no concerned or defining role.

Now if the identity of this person is dependent upon his identity as an agent at all, then there does not appear to be a *continuing subject* here either. As Frankfurt (2004: 23) says of such a person in a separate article, "Even if it could meaningfully be said of such a person that he had a will, it could hardly be said of him that his will was genuinely his own."

This idea of caring as critical to being one and the same person over time is underscored by the final reason in support of the importance of caring. This reason is the product of two distinct facts we have examined previously. First, there are no rationally warranted criteria for establishing anything as inherently or objectively important. Second, we have seen that caring is essential in providing us with final ends, saving us from boredom, and maintaining our identity as a continuing subject over time. The result:

> The significance to us of *caring* is thus more basic than the importance to us of *what* we care about. . . . [T]he value to us of the fact that we care about various things does not derive simply from the value or the suitability of the objects about which we care. Caring is important to us for its own sake, insofar as it is the

> indispensably foundational activity through which we provide continuity and coherence to our volitional lives. Regardless of whether its objects are appropriate, our caring about things possesses for us an inherent value by virtue of its essential role in making us the distinctive kind of creatures that we are. (Frankfurt 1999: 162-63)

IX. VOLITIONAL NECESSITY

The concept of volitional necessity is also related to personal identity, but before that connection is explained we need to explain the concept itself and how it completes the discussion of what we should care about. In his discussion of this topic, Frankfurt (2004: 56) claims that caring entails not only that we regard the objects of our caring as valuable in themselves, but also that "we have no choice but to adopt those objects as our final ends." While this claim will strike some as threatening to both autonomy and personal identity, it is actually indispensable to achieving both.

Recall that the question of what we should care about could not be coherently asked until the question of what we actually do care about was answered. The reason was that the answer to the factual question of what we do care about is the only metaphysically plausible, nonarbitrary basis upon which we could inquire what else we *should* care about. Now suppose that what we actually do care about is in fact entirely up to us and that it is not constrained in any way. In other words, suppose we could have whatever will we wanted—and thus care about whatever we wanted—simply by deciding to do so (remembering, of course, that caring is a fact about the will). How could the question of what to care about be answered if the basis for the answer

must be both nonarbitrary and entirely impartial (i.e., not taking into account any volitional predisposition)? This question cannot be answered. The answer must either be arbitrary or take into account things about which we *cannot help caring* (Frankfurt 1999: 93). This is because even raising the question of whether to keep your will as it is or to change it is to suspend the authority of any antecedent volitional state that could have provided the basis for an answer (Frankfurt 1999: 93). In order for a person to have an appropriate (i.e., nonarbitrary) basis for determining his final ends, two things must obtain. First, there must be something that is antecedently important to him (Frankfurt 1999: 93). And, as we saw before, these can change as they depend on a variety of causally influential factors. Second, even though what we care about can change, it cannot be subject to our own immediate, voluntary control (because if it were we would immediately be faced with the problem described at the beginning of this paragraph) (Frankfurt 1999: 93). This does not mean that the person cannot change it himself, but it cannot be as simple as his just making up his mind one way or the other (Frankfurt 1999: 94).

The things that are antecedently important to us that we cannot help caring about encompass the concept of volitional necessity, which is an instrumental concept for understanding several aspects of personhood. Autonomy, for example, requires volitional necessity. The things we cannot help caring about provide us with nonarbitrary reasons for exercising our autonomy. "Unless a person makes choices within restrictions from which he cannot escape by merely choosing to do so, the notion of self-direction, of

autonomy, cannot find a grip" (Frankfurt 1999: 110). This is because any choice a person totally lacking in any fixed volitional points makes cannot be regarded as having originated in *his own* will. To the extent that such a person can decide upon anything at all, his choice will be necessarily arbitrary, thus lacking any personal significance or authority that could serve as the basis for true autonomy.

Volitional necessity also provides a basis for establishing our identities as active beings. Frankfurt claims it is by these necessities that "our individual identities are most fully expressed and defined" (Frankfurt 2004: 50). He elaborates on this idea in a separate article:

> The essential nature of a person is constituted by his necessary *personal* characteristics. These characteristics have to do particularly with his nature as a person, rather than with his nature as a human being or as a biological organism of a certain type. They are especially characteristics of his will. In speaking of the personal characteristics of someone's will, I do not mean to refer simply to the desires or impulses that move him. We attribute impulses, desires, and motives even to infants and animals, creatures that cannot properly be said either to be persons or to possess wills. The personal characteristics of someone's will are reflexive, or higher-order, volitional features. They pertain to a person's efforts to negotiate his own way among the various impulses and desires by which he is moved, as he undertakes to identify himself more intimately with some of his own psychic characteristics and to distance himself from others.
>
> To be a *person* entails evaluative attitudes (not necessarily based on moral considerations) toward oneself. A person is a creature prepared to endorse or repudiate the motives from which he acts and to organize the preferences and priorities by which his choices are ordered. He is disposed to consider whether what attracts him is actually important to him. Instead of responding unreflectively to whatever he happens to feel most strongly, he undertakes to guide his conduct in accordance with what he really cares about.
>
> To the extent that a person is constrained by volitional necessities, there are certain things that he cannot help willing or

> that he cannot bring himself to do. These necessities substantially affect the actual course and character of his life. But they affect not only what he does; they limit the possibilities that are open to his will, that is, they determine what he cannot will and what he cannot help willing. Now the character of a person's will constitutes what he most centrally is. Accordingly, the volitional necessities that bind a person identify what he cannot help being. They are in this respect analogues of the logical or conceptual necessities that define the essential nature of a triangle. Just as the essence of a triangle consists in what it must be, so the essential nature of a person consists in what he must will. The boundaries of his will define his shape as a person. (Frankfurt 1999: 113-14)

Before discussing volitional necessity further, it is worth keeping in mind that the central topic of this project is personal welfare. Specifically, it may be objected here that the necessity of the sort described (or perhaps necessity of any kind) is antithetical to the very notion of personal welfare. I have some deep-seated sympathy for this position. However, I think the fear of the notion of necessity in the area of personal welfare must be abandoned, as I will attempt to show in the remainder of this section and in the next section on free will. Moreover, this section has already demonstrated how volitional necessity makes genuine autonomy possible and how volitional necessity is necessary in defining us as persons, both of which may be tangentially related to the concept of personal welfare.[39]

The remaining aspects of volitional necessity that are worthy of note can best be illustrated by using an example. Martin Luther's famous declaration,

---

[39] Autonomy is widely considered to be at least an important condition for the enhancement of personal welfare. This gives a coherent account of autonomy and supports the intuition that autonomy and welfare are related. The issue of personal identity could also be thought to be tied to the issue of personal welfare because one may want to establish the identity of the person who is to be the subject of the personal welfare calculation.

"Here I stand; I can do no other," is probably the best historical example to examine for this purpose. How are we to understand Luther here? He is saying that it is not possible for him to do other than he is doing. The impossibility is clearly not a logical necessity. It is also not a causal necessity, as Luther is fully aware that he possesses both the *capacity* and the *power* to do otherwise (Frankfurt 1988: 86). If we take Luther at his word, the only way to understand him is to describe him as lacking the *will* to do otherwise.

Accordingly, we could say that Luther, and anyone else subject to volitional necessity, *must* act as he does (Frankfurt 1988: 86-87). But if a person must act as he does in this situation, how are we to distinguish between volitional necessity on the one hand, and compulsions, obsessions, and addictions on the other? Since we may experience each of these as being irresistible, this cannot be the difference. The difference is that when we succumb to addiction and the like, we do so *unwillingly* and experience these forces as *alien* to ourselves (Frankfurt 1999: 136). The explanation of this goes back to the issue of identification; it is because we do not identify with obsessions and compulsions that we experience them in this way. When we do not identify with these irresistible forces, we do not want these motives to be ones that move us, and we deprive them of all authority. Considered strictly in themselves, this type of psychic raw material can only move us through sheer brute force (Frankfurt 1999: 137).

Volitional necessity is different. After all, there is no reason why irresistible forces must conflict with the desires by which we would prefer to be

moved (Frankfurt 1999: 136). Volitional necessity, as Frankfurt uses the term, describes the irresistible forces we identify with, endorse, and are pleased to have move us to action, even though we could not do otherwise. Accordingly, as in Luther's case, he does not accede to the manner in which he is compelled to act because he lacks sufficient strength of will to defeat it. "He accedes to it because he is *unwilling* to oppose it and because, furthermore, his unwillingness is *itself* something which he is unwilling to alter" (Frankfurt 1988: 87).

And it is precisely through identifying ourselves with the irresistible forces of volitional necessity and wanting to be moved by them that we experience these forces as liberating (Frankfurt 2004: 64). This is because the volitional necessity that binds the will puts an end to any indecisiveness concerning what we are to care about (Frankfurt 2004: 65). Our final ends are settled for us, we know what we are to care about, and we care about caring about those things. For these reasons, Frankfurt (2004: 64) likens the commanding necessity of volitional necessity to the commanding necessity of reason, in that "neither entails for us any sense of impotence or restriction." The reason is that both volitional and rational necessity eliminate uncertainty. They thereby lessen or remove self-doubt. In the case of reason, necessity tells us what *must* be the case and thus removes any doubts we might have concerning what to believe. In the case of the will, necessity tells us what we must care about and puts an end, at least for a time, to our indecisiveness in this area (Frankfurt 2004: 65). In other words, volitional necessity is liberating

rather than coercive because it may constrain the person to do what he really wants to do.

X. Free Will

The final topic in this chapter concerning the theoretical foundation of my theory of personal welfare is Frankfurt's theory of free will. The sole purpose of covering this theory is to explore the relationship, if any, between free will and personal welfare. The reason to inquire into this relationship involves a basic intuition concerning the value of free will in personal welfare calculations. Specifically, it seems reasonable to suppose that the exercise of free will would positively impact personal welfare, other things being equal. In other words, the idea that free will *does not* and *could not* have any effect at all on how well a life goes for the person who lives it may strike many as being counter-intuitive.

However, the devil is always in the details. There are several significant obstacles to showing how free will and personal welfare are related. Grounding the intuition concerning this relationship requires an account of free will and an account of personal welfare, ensuring not only that these accounts are internally coherent, but that they are coherent considered as a whole. Moreover, it requires an account of personal welfare that does not simply assign an *arbitrary* value to the effect of free will on the calculation—a problem that, as we will see later, is the most intractable one of the bunch for competing theories that may attempt to incorporate this basic intuition.

Each of these problems I have outlined comes with its own subset of problems. In the remainder of this section, I will limit the scope to giving an account of free will and the problems associated with doing so. To this end, it will be helpful to think about the requirements for a theory of free will along with other non-essential, yet desirable, features of any such theory. First and foremost, a theory of free will should be *coherent* and *plausible*. Although this hardly seems worth mentioning, the number of theories of free will on offer that are either incoherent or based on highly questionable metaphysics—or both—suggests otherwise. Second, a theory of free will should make it clear why we should care about it at all. Free will is commonly supposed to be humankind's most prized attribute (along with rationality), and a theory of free will should help us to understand why this is the case. Third, as it is also commonly supposed that animals do not possess free will, our theory should give us some reason to think that I do have free will, while the squirrels outside my window do not. Finally, given the content of the previous section on volitional necessity, the theory of free will presented here will need to explain how free will and volitional necessity are compatible.

The beginning is often a good place to begin, so let us start with Frankfurt's account of free will. In order to understand his account properly, we need to review some of Frankfurt's account of personhood. Persons, according to Frankfurt (1988: 16), are all and only those creatures that have second-order volitions. A second-order volition is a second-order desire for a certain first-order desire to be his will (Frankfurt 1988: 16). A person's will,

then, is identical to one or more of his *effective* first-order desires, which are ones that move (or will or would move) a person all the way to action (Frankfurt 1988: 14). Accordingly, to "identify an agent's will is either to identify the desire (or desires) by which he is motivated in some action he performs or to identify the desire (or desires) by which he will or would be motivated when or if he acts" (Frankfurt 1988: 14).

After providing this account of personhood, Frankfurt notes the close relationship between the capacity for forming higher-order volitions and the capacity for free will. These are so closely connected that, Frankfurt (1988: 19) states, the concept of a person could "also be construed as the concept of a type of entity for whom freedom of its will may be a problem." This is because freedom of the will can only be achieved through the formation of higher-order volitions, which, as we have just seen, are limited to persons.

Of course, just how a person's higher-order volitions and his will combine to produce free will needs to be explained. To do so, Frankfurt (1988: 20) makes an analogy between *freedom of action* and freedom of the will. Frankfurt (1988: 20) defines freedom of action as, "roughly, . . . the freedom to do what one wants to do." And although he thinks freedom of action and freedom of the will are analogous, Frankfurt (1988: 20) claims that freedom of action is both a distinct concept from freedom of the will and that freedom of action is neither a necessary nor a sufficient condition for freedom of the will. It is not a sufficient condition because we recognize that the squirrels outside my window enjoy freedom of action (i.e., they are free to run in whatever

direction they please), yet they do not enjoy freedom of the will (Frankfurt 1988: 20). It is also not a necessary condition. While it is often true that a person who is *aware* that he lacks freedom of action may feel the effect of this in the desires that comprise his will, which in turn will limit the range of decisions he is able to make, this is not the case for a person who is *unaware* that his freedom of action has been limited. Consider the case of a person who is either temporarily or permanently paralyzed, yet is unaware that this misfortune has befallen him. In such cases, the person's *will* is as free as it was before even though the desires and determination of his will that were transparently translated into action no longer have the same effect.

Accordingly, when we inquire about the freedom of a person's will, we are not asking if the person possesses freedom of action (i.e., whether he is in position to translate his first-order desires into action) (Frankfurt 1988: 20). Rather, as we can see from the paralysis example, we are inquiring into the person's desires themselves when we address the topic of free will, and it is here that the analogy between free action and free will proves useful. If freedom of action is roughly the freedom to do what one wants to do, then we can think of freedom of the will as, roughly, the freedom to want what one wants to want (Frankfurt 1988: 20). "More precisely, it means that he is free to will what he wants to will, or to have the will he wants" (Frankfurt 1988: 20). In other words, freedom of action concerns whether it is the action a person wants to perform and freedom of the will concerns whether it is the will the person wants to have (Frankfurt 1988: 20).

How, then, are we to understand a person having the will he wants in terms of the concepts relating to the will that we have already deployed? A person exercises freedom of the will by securing the conformity of his will to his higher-order volitions (Frankfurt 1988: 20). Therefore, there are two distinct ways in which a person may experience a lack of free will. The first way is obvious—when there is a discrepancy between his will and his higher-order volitions (Frankfurt 1988: 20). The second way is much less obvious and calls attention to the fact that the *person* must secure the conformity in question. A person also lacks free will if the person is aware that, although his will and his higher-order volitions are properly aligned, the coincidence of these two things "is *not his own doing* but only a happy chance" (Frankfurt 1988: 20).

To illustrate, let us examine the cases of three addicts who all lack free will, albeit for different reasons. Each of the three is physiologically addicted, and each has access to his drug of choice—glue. The first is an unwilling addict because his second-order volition is for his first-order desire not to sniff glue to be effective. The unwilling addict does not have free will with respect to this act because the will he has is not the will he wants. The second is a wanton addict because he either does not or cannot care which of his competing first-order desires wins out. He does not have the will he wants or the will he does not want. Since he is not a person (i.e., he is a creature without second-order volitions), freedom of the will cannot be a problem for him. "He lacks it, so to speak, by default" (Frankfurt 1988: 21). The third is a willing addict in that he has a second-order volition for his first-order desire to

sniff glue to be effective. Even though his will conforms to his second-order volition, he lacks free will. This is because the conformity *is not his own doing*. He is not free to have the will he wants because his first-order desire to sniff glue will be his will *regardless* of whether he wants this to be the case or not (Frankfurt 1988: 24-25).

What Frankfurt has supplied us with—a coherent and plausible account of free will—is no small feat. To appreciate this achievement, compare Frankfurt's account to one put forward by Roderick Chisholm. Every free action, on Chisholm's account, is a literal miracle (Frankfurt 1988: 23). Free acts are the outcome of a series of physical causes, but some event in the series, "and presumably one of those that took place within the brain, was caused by the agent and not by any other events" (Chisholm 1966: 18). This account elevates us to quite a lofty status by attributing to us "a prerogative which some would attribute only to God: each of us, when we act, is a prime mover unmoved" (Chisholm 1966: 23).

While this account appears to be coherent, it is far from plausible. The amazing metaphysics required to support this account of free will would only be rivaled by the amazing nature of the corresponding epistemology. However, Chisholm is by no means alone here. In fact, there is not a single satisfactory account of libertarian free will yet on offer.[40] And as the other group of

---

[40] As Michael Tooley (2010: § 7.2) notes:

> One problem with an appeal to libertarian free will is that no satisfactory account of the concept of libertarian free will is yet available. Thus, while the requirement that, in order to be free in

incompatibilists—hard determinists—denies the existence of free will altogether, the only satisfactory account of free will will be a compatibilist account. Moreover, of the three main types of compatibilist accounts of free will, the type developed by Frankfurt is an example of the most common, and most plausible, type.[41]

Frankfurt's account also provides us with an account of why free will is thought to be desirable. As Tooley (2010: §7.2) notes, why free will, on a libertarian account, should be thought valuable "is far from clear." Consider again the Chisholm account. As Frankfurt (1988: 23) notes:

> But why, in any case, should anyone *care* whether he can interrupt the natural order of causes in the way Chisholm describes? Chisholm offers no reason for believing that there is a discernible difference between the experience of a man who miraculously initiates a series of causes when he moves his hand and a man who moves his hand without any such breach of the

---

> the libertarian sense, an action not have any cause that lies outside the agent is unproblematic, this is obviously not a sufficient condition, since this condition would be satisfied if the behavior in question was caused by random events within the agent. So one needs to add that the agent is, in some sense, the cause of the action. But how is the causation in question to be understood? Present accounts of the metaphysics of causation typically treat causes as states of affairs. If, however, one adopts such an approach, then it seems that all that one has when an action is freely done, in the libertarian sense, is that there is some uncaused mental state of the agent that causally gives rise to the relevant behavior, and why freedom, thus understood, should be thought valuable, is far from clear.
>
> The alternative is to shift from event-causation to what is referred to as 'agent-causation'. But then the problem is that there is no satisfactory account of agent-causation.

[41] A claim I do not have the space to defend here. For a discussion of free will and the three compatibilist positions, see Derk Pereboom's (2001: 571-74) account in the *Encyclopedia of Ethics*.

> normal causal sequence. There appears to be no concrete basis for preferring to be involved in the one state of affairs rather than in the other.

Frankfurt's account, on the other hand, does provide a reason for desiring free will. The experience of exercising free will entails the satisfaction of higher-order desires that signify that his will is his own (Frankfurt 1988: 22). The frustrations of such desires suffered by a person experiencing the lack of free will are those of a person who is being violated by forces that he has attempted to alienate from himself (Frankfurt 1988: 22).

In addition, Frankfurt's account provides us with a strong basis for doubting that most nonhuman animals have free will. If an animal lacks the requisite reflexive capacity, then it would also lack free will. In other words, if a creature lacks the capacity to *know* that it has a will, then it would also seem to lack an ability to exercise the control of its will that could serve as the basis for ascribing free will to that creature. As Frankfurt (1988: 23) notes, "Chisholm says nothing that makes it seem less likely that a rabbit performs a miracle when it moves its leg than that a man does so when he moves his hand."

Before revisiting the issue of volitional necessity, this time examining its relationship to free will, let us take stock of where we are in terms of the amount of freedom it is possible for us to enjoy:

> Suppose now that someone is performing an action that he wants to perform; and suppose further that his motive in performing this action is a motive by which he truly wants to be motivated. This person is in no way unwilling or indifferent either with respect to what he is doing or with respect to the desire that moves him to do it. In other words, neither the action nor the

> desire that motivates it is imposed upon him against his will or without his acceptance. With respect neither to the one nor to the other is he merely a passive bystander or a victim.
>
> Under these conditions, I believe, the person is enjoying as much freedom as it is reasonable for us to desire. Indeed, it seems to me that he is enjoying as much freedom as it is possible for us to conceive. This is as close to freedom of the will as finite beings, who do not create themselves, can intelligibly hope to come. (Frankfurt 2004: 20)

There are two ideas here that bear highlighting, the first of which is that *we do not create ourselves*. This seems fairly uncontroversial, but the implications of this are often thought to be very controversial. This raises the issue of causal determinism and the view that free will and determinism are incompatible. Determinism does not threaten our *freedom*, but our *power*. We human persons are not omnipotent:

> As finite creatures, we are unavoidably subject to forces other than our own. What we do is, at least in part, the outcome of causes that stretch back indefinitely into the past. This means that we cannot design our lives from scratch, entirely unconstrained by any antecedent and external conditions. However, there is no reason why a sequence of causes, outside our control and indifferent to our interests and wishes, might not happen to lead to the harmonious volitional structure in which the free will of a person consists. That same structural unanimity might also conceivably be an outcome of equally blind chance. Whether causal determinism is true or whether it is false, then, the wills of at least some of us may at least sometimes be free. In fact, this freedom is clearly not at all uncommon. (Frankfurt 2006: 16)

All of this points to the second item of note in the last excerpt. This is the idea that when we act freely, the key is that we are not passive victims. There is no sense in which we are being violated or defeated or coerced. The essential element, then, is that we are autonomous when we are free (i.e., "whether we are active rather than passive in our motives and choices—whether, however

we acquire them, they are the motives and choices that we really want and are therefore in no way alien to us") (Frankfurt 2004: 20 n.5).

Finally, just as causal necessity does not threaten our freedom or autonomy, volitional necessity does not do so either. We can see this by thinking about what free will is—being free to have the will one wants. Volitional necessity, when present, simply ensures that the person does not *want* another first-order desire to be his will. It is in these cases that we can feel our autonomy is being enhanced by eliminating any uncertainty as to what we should care about. Contrast this with another type of necessity that does threaten our freedom. When the necessity comes from our first-order desires, and not our higher-order volitions, then we experience a lack of free will. As with each of our three physiologically addicted glue addicts, they are not free to have the will they want. At worst, they will be dissatisfied with the motive that moves them, thus ensuring that they are passive victims of forces that coerce and defeat them. At best, they will recognize that the will that moves them is not of their own doing, but is merely a fortuitous coincidence. Accordingly, volitional necessity, like causal necessity, does not threaten our autonomy, but rather enhances it.

In the next chapter we will put the final pieces of our theory in place before—in the last chapter—setting out the theory in detail and seeing it in action.

CHAPTER FOUR: PUTTING THE FINAL PIECES IN PLACE

With the Frankfurtian foundations solidly in place, we are well on our way to setting out a complete theory of personal welfare—a task that will be accomplished in the next chapter. However, we do not yet have all the tools we will need at our disposal. That is the goal of this chapter. While it should be clear that the next chapter will contain a desire-satisfaction theory based on Frankfurt's conception of the will, there is more to the story.

The rest of the story will be filled in by surveying the desire-satisfaction literature with an eye to separating the wheat from the chaff. The more detailed and precise explanation for the content and structure of this chapter is as follows. As discussed previously, desire-satisfaction theories that in no way limit the desire satisfactions that are claimed to increase personal welfare (i.e., *unrestricted* desire satisfactionism) are not plausible. Critics have suggested that there are many types of desires the satisfaction of which, at best, do not make a life go better for the person who lives it, or, at worst, actually decrease personal welfare. Every major desire-satisfaction theorist has taken some subset of these objections to heart, and, in turn, each of them presents a version of a *restricted* desire-satisfaction theory. As I intend to follow this same general strategy, this chapter will contain an examination of the most common types—as well as the most promising types—of restrictions. The goal is to come up with a list of restrictions and other considerations that must be taken into account in constructing the theory to be presented in the next chapter.

I. Idealized Desires

As mentioned above, both critics and proponents of desire theory claim that the satisfaction of some desires either has no effect on personal welfare or even makes the person's life go worse. Heathwood presents a nice sampling of these so-called *defective desires*.[1] Weak-willed Willie, who has an irrational fear of dentists, wants to avoid the dentist more than he wants relief from his suffering (Heathwood: 6). Unimaginative Maggie fails to adequately appreciate how bad the paparazzi will make her life when she satisfies her desire of becoming a pop star (Heathwood: 6). Ignorant Ignacio wants to drink water from a river that, unbeknownst to him, is poisonous and will make him sick for months (Heathwood: 6). It is these sorts of desires, among others, that have persuaded some that desire satisfactionism must be abandoned entirely and others to adopt a restricted form of the theory.

The strategy most commonly employed to deal with the problem of defective desires is to restrict desire satisfactionism to desires that have been *idealized* in some way.[2] Ideal desires are those desires that have supposedly been rendered free from all defects by going through the process of idealization. In other words, ideal desires are those desires you would have if your desires were either subjected to or produced by *Process X*. Process X has taken several

---

[1] Brandt (1979: 88) notes that the idea of defective desires goes back to the Greeks, who considered "intense desire for reputation, or power, or wealth" to fall into this category.

[2] For a nearly exhaustive list of the ways this can be achieved, see Fehige & Wessel's *Preferences* (1998: xxv-xxvi).

slightly different forms, but it appears to have its modern origins in Sidgwick.[3] Sidgwick (1907: 711-12) explores the claim that a person's good is "what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were adequately foreseen and adequately realized in imagination at the present point in time." Rawls (1999: 366) then uses Sidgwick's idea in formulating his notion of deliberative rationality, claiming the rational plan for a person

> is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires.

This statement contains the seeds for all the idealized versions of desire satisfactionism. This is because Rawls's (1999: 366) deliberative rationality requires that there be "no errors of calculation or reasoning," requires that "facts are correctly assessed," requires that the agent possess "full information," requires that "the agent is under no misconception as to what he really wants," and presumes that "the agent's knowledge of his situation and the consequences of carrying out each plan . . . [is] accurate and complete." Rawls (1999: 366) claims that his version of Process X "is the objectively rational plan for him and determines his real good."

Brandt subsequently uses a very similar idea in discussing his notion of rational desires. Brandt (1979: 88) claims that, after exposing a person to

---

[3] Another possibility for the modern origins of these theories is Mill's (2006: 321-22) competent-judges test as a means for determining the quality of different pleasure.

cognitive psychotherapy (his version of Process X), there are some desires "which a fully rational person would not experience." Cognitive psychotherapy, for Brandt (1979: 113), is the "whole process of confronting desires with relevant information, by repeatedly representing it in an ideally vivid way, and at an appropriate time." Similar views have also been accepted elsewhere.[4]

On the surface, the appeal to desires that have been idealized in some way as a solution to the problem of defective desires seems very promising. Take defective desires, add just the right amount of information and/or error-free processing power, and—Shazam!—the resulting desires, no longer defective, are ensured of increasing personal welfare upon satisfaction. Of course, it is not that easy. The devil is always in the details, and nowhere is that more true than here. Several objections have been leveled against both the concept of ideal desires and the process used to produce those desires. While the following list is not exhaustive, it does include a brief summary of most, if not all, of the more popular objections:

- *Statements about what a person would desire after undergoing Process X lack determinate meaning and/or a truth value.* Theories making use of idealized desires rely on counterfactual statements about what a person would desire if she underwent Process X. However, it has been argued that statements of this sort lack a determinate meaning and truth value. For example, J. David Velleman (1988: 365) argues that fully representing

[4] See, e.g., Hare (1981: 214-16), Darwall (1983: 85-100), Griffin (1986: 26-31), Lewis (1989: 121-26), and Overvold (1982: 188).

information to yourself does not refer to any determinate state of affairs because "the same facts can be represented in many different ways with different motivational consequences." Connie Rosati (1995: 309) makes a similar point in discussing the problem of "experiential ordering." Roughly, the idea here is that facts must be presented in some sort of order, the particular order may affect our final judgment or attitude about those facts, and there is no single "correct" way to present those facts. For additional statements of this type of objection, see, e.g., Carson (2000: 226-29), Adams (1999: 86), Loeb (1995: 4), and Griffin (1986: 16).

- *Some defective desires can survive Process X.* Gibbard (1990: 19-20) provides an example of a defective desire that could survive Brandt's cognitive psychotherapy. He asks us to imagine a person who is so obsessed with germs that he washes his hands several times each hour. However, even after undergoing the process of being vividly and repeatedly made aware of the relevant facts about the relatively low risk that germs pose and the opportunity costs associated with the frequent washings, he still does not want to have those "creepy-crawly things" on his hands and does not "want to be a person who would be willing to tolerate them on his hands" (Gibbard 1990: 20). For additional statements of this objection, see, e.g., Adams (1999: 87), Loeb (1995: 9-11), Sumner (1996: 130-32), Heathwood (2006: 545), and Carson (2000: 226).

- *Process X could fundamentally change the person.* This idea is stated in a number of ways, but the underlying idea is that the entity that undergoes

Process X is radically different from the entity that existed before the process. Rosati (1995: 310-11) is concerned that the person exposed to this process will not still be *you* in any meaningful sense.[5] Loeb (1995: 13-14) states that it would be impossible for "ordinary people" to undergo the required change in their intellectual and imaginative powers—so much so that the beings undergoing this would have to be "vastly different from ordinary people." Carson (2000: 229) takes this line of thought to its logical conclusion by questioning "whether it makes sense to say that someone who is fully informed is a *human being*." This objection is pursued extensively in Sobel's (1994: 794) article wherein he argues that the notion of a fully informed self is a "chimera" because all the available lives that need to be considered cannot exist in one's consciousness. The problem with the change in the person from the actual to the ideal, if this is even possible, is that "it would be surprising if the well-being of the two of us . . . consisted in the same things" (Sobel 1994: 793).[6] Indeed, it would seem counterintuitive at best and dogmatic at worst to insist that the good for the person who underwent Process X *must always be identical* to the good of the person who began the process.[7]

---

[5] A similar objection is presented by Adams (1999: 86).

[6] This point is also made, albeit in slightly different forms, by Rosati (1995: 299) and Loeb (1995: 16).

[7] If you are still unpersuaded on this point, consider the case of Dullard Dan. Dullard Dan, a 40-year-old junior high school graduate, loves Pabst Blue Ribbon, TV dinners, and NASCAR. Now we take Dan and send him through Process X, which can be properly characterized as a kind of turbo-charged education/souped-up broadening-of-horizons project. Then we have to answer

- *Process X could change the person for the worse.* This objection puts pressure on the idea that the process of idealization will, at the very least, leave a person in a better condition than she was before. When Process X requires full information, Carson (2000: 229-30) questions, for example, whether a person would lose her sanity if she were vividly aware, in a very detailed fashion, of all the suffering caused by World War II. Similarly, Loeb (1995: 19-20) wonders whether the idealized version of oneself might be too depressed to care about the actual person's welfare if that idealized version had been exposed to every kind of pain imaginable (e.g., burned to death, drowned, eaten alive, etc.). Gibbard (1990: 20-21) worries that a person who is vividly aware of what people's innards are like might never want to eat around people again or that an honest civil servant might start taking bribes if he became aware of the nice things this could lead to. Arneson (1999: 133-34) believes this to be the most serious objection to theories based on idealized desires, and Rosati (1995: 312-13) and Velleman (1988: 360) offer similar objections.

- *Objective-list theories may be masquerading as the problem of defective desires.* Many of the theories based on idealized desires seem to be little more than the objective-list theory wolf dressed in desire theory sheep's clothing. Rather than mounting a frontal assault on desire theory by providing a list of

---

the question: Is the good the same for the Dan who entered the process and the Dan who exited the process? Of course it can be claimed—perhaps even plausibly—that the good for these two might be the same, but that is not strong enough for the proponents of this version of desire theory. The conclusion that they must reach is that the good for these versions of Dan *must always be identical*. This is not plausible.

things that make a person's life go better whether he wants them or not, the cunning objective-list theorist secretly parachutes behind enemy limes, dons the uniform of a quietly slain desire satisfaction theorist, and then discovers some "defective" desires. Needless to say, the proposed remedy to this problem turns out to be to substitute the desires you *would* have if you implement Process X for those desires that you actually *do* have. While this is a much better marketing plan for the objective-list theorist, the result is the same: a list of things that will make your life go better no matter what your actual desires might be. Sen (1985: 191-92) presents the objection this way:

> To ask what one would desire in unspecified circumstances—abstracting from the concreteness of *everyone's* life—is to misunderstand the nature of desire and its place in human life. Of course, we can *pretend* to answer this question. Since all this is imaginary anyway, we need not live in fear of being proved wrong. This can be done by making—explicitly or by implication—some simple assumption, e.g., that our desires would be in line with what Scanlon has called "an objective criterion" of well-being, appealing to a certain "consensus" of values about the content of well-being. But if that is what we are going to do, we could just as easily have started from the objective criterion itself, and "founded" it on the consensus of values on well-being, rather than having the imaginary exercise of *counterfactual desiring*.

Feldman (2004: 17) makes this form of the objection explicit, while Sobel (1994: 795) ("the purportedly naturalistically described informed viewpoint invokes unreduced normative notions") and Adams (1999: 86) (idealized versions of desire satisfaction guilty of paternalism) offer the same objection using more euphemistic terms.

Does this list of objections render idealized versions of desire satisfactionism implausible beyond repair? The answer to this question can

best be assessed by examining the two most plausible versions. First, Peter Railton (2003: 54) proposes a version that purportedly avoids the objections listed above:

> The proposal I would make, then, is the following: an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality. The wants in question, then, are wants regarding what he would seek were he to assume the place of his actual, incompletely informed and imperfectly rational self, taking into account the changes that self is capable of, the costs of those changes, and so on.

Carson (2000: 244), in turn, provides us with a similar, albeit turbo-charged, version of the theory wherein God is substituted for the idealized version of you, and we look to what God would prefer you to prefer.[8] Carson's version trades in the advantage of Railton's theory—that the idealized entity is actually "you"—for the advantages of addressing the concerns about humans not having the capacity to undergo the required process, about the process changing the person for the worse, and about statements concerning what the idealized entity would want lacking truth values.

---

[8] While Carson actually specifies that we are to consider what a *loving* God would prefer you to prefer, this aspect of his theory can be safely ignored. This is because Carson's use of a loving God is either circular or fails to accomplish its goal. It is quite obviously circular if *loving* entails increasing the welfare of the person in question (as most ordinary uses of loving, such as in a parent-child relationship, seem to do). If it does not incorporate notions of increasing welfare, then it is very unclear how adding the qualifier of "loving" to God is supposed to *ensure* that the preferences of this entity will increase a person's welfare in a way that just plain ol' God's preferences would not (leaving aside, of course, that the mere mention of divine intervention with regard to any problem is supposed to remedy that problem without the need for questioning how it is supposed to occur).

However, it is not necessary to assess the relative merits of these two competing theories because, even if they avoid all the objections listed above,[9] there are at least two additional decisive objections. First, both of these theories have to allow for the possibility that either God or your idealized self does not like you.[10] A little imagination is required to understand this possibility, but not much. Imagine how you or I might be perceived by God or your idealized self (which seems to be nothing more than a Mini-Me version of God). We would be perceived as being physically weak, stupid, morally iffy at best, weak-willed, and probably just generally silly. Now what impact does this have on what these beings would prefer me to prefer? It appears as though *nothing* is entailed in this scenario, and therein lies the problem. God or Mini-Me God may find me too pathetic to contemplate and therefore may assign a subordinate to pick preferences randomly out of a hat that are supposed to be what he wants me to want. God or Mini-Me God may hate me in the way that some humans hate rats, spiders, snakes, or some parasites and make it his mission to prefer that I prefer things that will make my life go as poorly for me as possible. Of course, God or Mini-Me God may take a shine to my perfect

[9] It is highly unlikely that either Railton's or Carson's theory do avoid all the earlier objections. For example, it is unlikely that both theories avoid the objection that some defective desires can survive idealization and the objection that idealization is just a disguise for objective-list theory, as these are the Scylla and Charybdis of idealized theories. If a theory is not actually an objective-list theory, then some defective desires are bound to survive idealization. Alternatively, if no defective desires survive idealization, then it is likely an objective-list theory in disguise.

[10] *Fight Club*'s Tyler Durden states this objection with gusto: "You have to consider the possibility that God does not like you. He never wanted you. In all probability, he hates you. This is not the worst thing that can happen."

instantiation of patheticness and prefer me to prefer things that will make my life go swimmingly. The problem for these theories is that there does not appear to be any non-circular way to ensure that this third possibility is the case and that the first two (or any of the other myriad of possibilities) never occur.

The second objection is one we have seen before. The reason it is applicable here is the same reason it was applicable before, and it pertains to what all idealized desire theories are at their core. All of these theories claim that a person's life goes better for her if desires that she *would have if some counterfactual event were to occur* are satisfied. Now, of course, a person may actually have some of the desires that the idealized theory in question would deem welfare-enhancing, but this is a contingent matter, as nothing in these theories ensures that any person will have even one of these desires. In other words, these idealized desire theories claim that a person's life goes better for her if she has desires—*that she actually does not have*—satisfied. More straightforwardly false claims are hard to find in modern philosophy. And it is false because it violates our Internalist Principle' (IP'): The value of a life (or part of a life) for the one who lives it is determined to a significant degree by what the person in question cares about. Another way of stating this objection is that "unwanted satisfactions of merely ideal desires are not . . . necessarily intrinsically good for a person," (Heathwood 2006: 545)—an objection that has

been repeatedly accepted in the literature.[11] To see this point more vividly, imagine a person who (1) gets all of the things that would be recommended by the idealized desire satisfaction theory in operation, and (2) absolutely loathes each and every confounded thing recommended by the theory when she gets it. The theory would claim that her life went very well for her. This is a claim that simply cannot be taken seriously.

II. First Fix: Future Desires

Although the appeal to ideal desires is unsuccessful, there are lessons to be learned. The first lesson, which is really just a reinforcement of a lesson we have already learned, is that the right theory of personal welfare must take our actual desires into account. Constantly telling a person throughout his life what desires he should have and then satisfying these desires for him, without more, does him no favors.

The second lesson is less obvious. Recall that the concern that prompted the move to idealization was various types of "defective" desires. Now let us assume that ideal desire-satisfaction theories are not just incognito objective-list theories. This will mean that all objects are possible objects of desire for purposes of increasing personal welfare. If this is the case, then what are the remaining possible ways in which a desire may actually be defective relative to welfare? One way is that the satisfaction of the desire in question may lead to an increase in the ratio of desire frustrations to desire satisfactions than would

[11] See, e.g., Heathwood (2006: 545), Griffin (1986: 11-12), Sobel (1994: 792-93), and Feldman (2004: 17).

have been the case had the desire in question been frustrated instead. In other words, the desire is defective because, due to the relative mix of *future* desire satisfactions and frustrations, the satisfaction of the desire *leads to* a decrease in personal welfare when compared to what would have happened if the desire had not been satisfied.

A closer examination of the defective desires mentioned above will help to make this point clear. Filling in the details for each will make the defect explicit as well as demonstrate how an appeal to actual future desires, rather than idealized desires, will yield the correct answer in each case. Recall Heathwood's examples of defective desires. Weak-willed Willie has an intense fear of dentists that prevents him from going to the dentist to seek relief from a tooth problem. The intuitive appeal of this case as an example of a defective desire stems from the assumptions we make about Willie and his two possible paths. Let us call Path One the path in which Willie does not go to the dentist to have the cause of his discomfort treated, and Path Two the one in which he does go to the dentist. The intuition here is that, on Path One, the suffering caused by his tooth will eventually lead to a desire frustration total that will outweigh the desire satisfaction provided by his not seeing the dentist. Accordingly, Path Two will lead to a better personal-welfare outcome for Willie. If this is the correct assessment of Willie's desires on each path, then the desire to avoid the dentist is defective. Taking notice of the fact that it is not at all clear that this defective desire could not survive idealization, what is clear is

that an appeal to Willie's actual future desires accurately tells us that Willie is worse off for not going to the dentist.

The same holds true for both Unimaginative Maggie and Ignorant Ignacio. The guiding intuition in the case of Unimaginative Maggie is that the balance of desire satisfactions and frustrations in the life in which she is a pop star is less favorable than in the life in which she is not a pop star due to her failing to properly foresee that, although her desire to be a pop star is satisfied, there will be associated desire frustrations relating to the paparazzi, being stalked, losing privacy, etc. Ignorant Ignacio has a similar problem. The intuitive assumption here is that when Ignacio drinks the water that is—unbeknownst to him—tainted, the desire satisfaction he gets from drinking the water will be more than outweighed by the myriad of desire frustrations that will befall him as a result of this drink. If this is in fact the correct assessment of Maggie's and Ignacio's desires on each of their two potential paths, then these desires are defective as they relate to personal welfare, which can quite accurately be shown simply by appealing to their actual future desires.[12]

---

[12] Indeed, the appeal to actual future desires for a desire theorist seems almost unavoidable, as this scene from the movie *Charlie Wilson's War* nicely illustrates:

> GUST: Listen, not for nothing, but do you know the story about the Zen master and the little boy?
> CHARLIE: Oh, is this something from Nitsa, the Greek witch of Aquilippa, Pennsylvania?
> GUST: Yeah, as a matter of fact, it is. There was a little boy, and on his 14th birthday he gets a horse. And everybody in the village says, "How wonderful! The boy got a horse." And the Zen master

Now, of course, the right theory will not appeal to all of a person's actual desires since, as we have seen before, they may not be desires that the person cares about. However, the appeal to (the proper subset of) actual future desires seems to be such an obvious and elegant solution to the problem of defective desires that one may wonder how the messy and complicated versions of ideal desire theory even came to exist. Upon reflection, the course of this tedious jaunt down the wrong track can be easily imagined. It seems to have started with an intuition, or maybe just a plain ol' desire, that Sidgwick and Rawls shared regarding the potential usefulness of the theory. (I used to share it as well, and I suspect many other people do, too.) It can be seen in the way that both Sidgwick and Rawls formulate their ideas in this area. Sidgwick (1907: 711-12) considers the view that a person's good is "what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were adequately foreseen and adequately realized in imagination at the present point in time," and Rawls (1999: 366) says that the rational plan for a person "is the plan that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the

---

> says, "We'll see." Two years later, the boy falls off the horse, breaks his leg. And everybody in the village says, "How terrible!" And the Zen master says, "We'll see." Then a war breaks out, and all the young men have to go off and fight, except the boy can't 'cause his leg's all messed up. And everybody in the village says, "How wonderful!"
> CHARLIE: And the Zen master says, "We'll see."

The obvious moral to the story is that the satisfaction of any desire—almost no matter how good or bad it may seem at the time—can start a chain of events that leaves one with radically different appraisals at different times.

relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize his more fundamental desires." Notice that each formulation involves an agent looking forward in time and trying to ascertain what he should do from the standpoint of increasing his personal welfare. In short, Sidgwick and Rawls want their ideas to be *useful*. As in: Dear Agent, the best course of action for you can be figured out, more or less, if you get the facts about your options and then properly ponder those facts relative to your psychic makeup—now get crackin'! Philosophers coming after Sidgwick and Rawls, given their tremendous stature, simply picked up this trail and figured that the answer must be down at the end of the trail somewhere. The problem is that, as between a *useful* theory and a *right* theory, we (hopefully) obviously want the right one. And there is no guarantee (or even any *a priori* likelihood) that the right theory will either be as useful as some might like or even be useful at all. The right theory's appeal to actual future desires involving counterfactuals has just this feature relative to the more useful Sidgwick/Rawls approach. While this is another unfortunate fact about reality, it does nothing to undermine the truth of the theory.

III. Remote Desires

The problem of remote desires was touched on in Chapter Two and involves the fact that it is *possible* to desire anything at all, even if the existence of the object of the desire is *not possible*. Thus, it is at least conceptually possible to want *anything*, to want any conjunction of things all the way up to *all things* (including things that are merely possible and not

actual), and to want those "things" that are neither actual nor possible. As mentioned before, the problem lurking here is "that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one would take as touching one's well-being" (Griffin 1986: 17). The problem, in other words, is that there may be too much conceptual space between a desire and our personal welfare, thus rendering the desire too *remote* to be a factor in our well-being. Sumner (1996: 135), who refers to this issue as a "problem of scope," states that these "problems of scope can be regarded as an invitation to qualify the theory so as to contour desire-satisfaction better to well-being."

Before we accept the invitation Sumner mentions, it will be worthwhile to survey the various examples of alleged remote desires in the literature. The force of the objection from remote desires relies on at least three different intuitions concerning welfare. Sumner (1996: 125) provides a good example of the type of case that is supposed to produce the first of these intuitions: Sumner asks us to suppose that his brother, who suffers from a debilitating disease for which he cannot get adequate treatment at home, moves to Papua New Guinea where a promising new treatment is available. Sumner's brother is subsequently cured—thus satisfying Sumner's desire that this state of affairs occur—yet Sumner never knows of this because his brother broke off contact

with Sumner after moving abroad.[13] Sumner questions how a desire satisfaction like this can make him better off.

At first glance, examples like this seem to rely on the spatial proximity between the desires and the state of affairs in question. Parfit (1984: 494-95), for example, gives several scenarios that involve his being in "exile" and desiring outcomes that involve the children he left behind—spatially remote states of affairs. However, it is highly implausible that the distance between the desires and the relevant states of affairs is relevant in any way. To illustrate this, let us change Sumner's example so that he lives in one half of a duplex and his brother lives in the other half. One day, Sumner's brother tells him he has cancer, erects a massive wall over the common wall of the duplex so that Sumner never again sees his brother's comings and goings, and also, as in the previous example, cuts off contact with Sumner. Again, unbeknownst to Sumner, his brother is cured, which is what Sumner wants. Now, whose life is going better: Papua New Guinea Sumner or Duplex Sumner? One has to conclude that these Sumners are equal with respect to personal welfare unless one wants to introduce a spatial component into desire theory. This is not plausible. What is plausible to suggest is that the relevant factor in the Sumner and Parfit examples is that the states of affairs in question are remote, not in spatial terms, but in terms of what Sumner and Parfit are *aware of*.

---

[13] Parfit (1984: 494) gives a similar example involving a stranger instead of a brother—a substitution that clouds the specific kind of intuition we are supposed to be relying on in the case.

The second intuition driving the objection from remote desires can be seen in an example from Shelly Kagan (1998: 37). Kagan asks us to imagine that, as a fan of prime numbers, he wants the number of atoms in the universe to be a prime number. Kagan says it is "absurd" to believe that his life is going better if this desire is satisfied. We could take this to be another example of the intuition Sumner was driving at, but Kagan could also plausibly be driving at something else entirely. Perhaps Kagan is relying on the fact that it will strike most people as highly implausible that anyone could really find that the number of atoms in the universe is a fact that is important to him. In short, this state of affairs is remote from what we *care* about.

To appreciate the third driving intuition behind the objection from remote desires, a very brief recap of what standard versions of desire theory claim is in order. The theory claims that personal welfare is enhanced when (1) a person desires state of affairs P, and (2) state of affairs P obtains. So now suppose that when I am seven years old I want a gorilla named Davey to beat up the skateboard kids who pull on my underwear and, being the moral little kid I am, I want the gorilla to take his orders from The Talking Walnut so that it wouldn't be my bad thing. Lo and behold this comes to pass as a Festivus miracle! According to the theory, my life is going better than it was before this well-deserved beating came to pass. But is it really? The details I neglected to mentioned are that 75 years have passed, I had long ago totally forgotten about this desire, and the skateboard kids are now wheelchair invalids. The problem

here, or at least one of the problems, is that this desire satisfaction seems too remote *in time* to have any impact on my well-being.

A particularly persuasive subset of these desires that are too remote in time includes the desires of the dead. So if one were inclined to think that the beating in the last example did make my life go better for me, does it matter if I die sometime in the interval between the desire and its satisfaction? In other words, can my personal welfare be affected after I am dead? No. It seems wildly implausible to claim that my life is going better *for me* at any point after I am dead, as the following example should decisively demonstrate. Suppose I become a loopy narcissist immediately after defending my brilliant dissertation, and I want people from that day forward to approach me, genuflect, recite a few lines of my dissertation, and then yell "Hallelujah Hyde!" After years of this pure awesomeness, I form very strong desires (one for each person-day combination) that, upon my death, every living person on each day at noon for the rest of time stop what he or she is doing and perform this ritual with a blow-up doll of me instead of actual me. If these posthumous desire satisfactions and frustrations can affect my personal welfare, then the assessment of my welfare will likely look radically different 100 years after my death than it did the day I died. Although this is not necessarily the case (these posthumous satisfactions and frustrations may perfectly offset each other during the intervening 100 years), what is necessarily the case is that a person who only knows the final assessment of my welfare 100 years after my death will have no idea how to answer the question of how well my life went for

me *while I was alive*.[14] This is a serious defect.[15] Any theory with this implication should be rejected.[16]

IV. Second Fix: (True?) Belief

Just as the move to idealized desires had something to teach us about how the right theory should look, the objection from remote desires does as well. In particular, this section will address, mainly, the objection from desires that are remote from what we are aware of (illustrated in the last section by the examples of Sumner's sick brother and Parfit's exile). The guiding intuition here is that desire satisfactions or frustrations that one is not aware of do not affect one's personal welfare. The following hypothetical may help strengthen this intuition. Suppose one of Satan's helpers, Stan, is assigned the task of ensuring, with respect to all the welfare-relevant desires I have, that I never become aware of whether these desires are satisfied or frustrated (at least with respect to those desires that *could* be satisfied or frustrated without my awareness). If some version of desire satisfactionism is true (which should be

---

[14] Unless, of course, one holds that what goods one enjoys while one is alive is not affected by posthumous events. However, to hold such a view simply underlines the fact that this is *not* a view about how well a life goes *for* the one who lives it.

[15] "Let us discuss the changes in how well George Carlin's life is going *for him* since his death." If someone were to start a conversation with me using this line, I would, at the very least, have to question his command of the English language.

[16] For an endorsement of this position regarding posthumous desires, see, e.g., Overvold (1980: 108), Haslett (1990: 81), and Fuchs (1993: 215-20). For a rejection of this position, see, e.g., Carson (2000: 76-77) and Portmore (2007: 27).

well-settled by this point in the proceedings), then it seems as though Stan's efforts will wreak havoc on my personal welfare by making it very difficult for my life to go better or worse. Why do Stan's meddlings have this effect on my welfare?

To explain this fact, it may be useful to return to the distinction between the concepts of *good for the world* and *good for a subject*. It is plausible to claim that, when a desire is satisfied, this is good for the world. In other words, when a desire is satisfied, other things being equal, this makes the world a better place. Yet it is not clear how this is better *for the desirer*, because the desirer has no idea if the desire has been satisfied or frustrated.[17] We could explain how this in fact becomes good *for* the desirer if we add in the requirement that the desirer actually believe the desired state of affairs has obtained. Griffin (1986: 13) calls this the Experience Requirement and describes it as "the link between 'fulfillment of desire' and the requirement that the person in some way experience its fulfillment" if it is to make the person better off.[18]

Before turning to the question of whether something in addition to belief is required, a brief examination of what, exactly, must be believed is in order. The obvious first step here is to claim that, if one desires X, then one must just

---

[17] Recall that the technical definition of a desire satisfaction is merely that the desired state of affairs obtains—not that the desirer be aware that the state of affairs has obtained.

[18] Kagan (1992: 186) appears to be making a similar claim when he says that for something to genuinely benefit a person, "it must make a difference *in* the person."

simply believe X has obtained. This is the approach favored by Heathwood (17): "Thus I propose that we require only that the subjects believe that the proposition desired is true . . . ." Although Heathwood (17) recognizes that the incorporation of the Experience Requirement "makes a serious break with traditional desire theory," the change he proposes marks a break that is *too* serious. While the objection from remote desires taught us that the constituents of our welfare must enter our experience, we would be wise not to completely abandon the general approach of desire theory that has served us so well to this point. Accordingly, the problem with the Heathwood approach is not the incorporation of the Experience Requirement, it is the fact that a major requirement of desire satisfactionism has been jettisoned in the process. Recall that desire theory claims that your life is going better for you once your desire *has been satisfied*. Not only does Heathwood's theory not require that your desire actually be satisfied in order to enhance your personal welfare (more on this below), *you also do not need to believe that it has been satisfied.* An example of this is Heathwood's (32) claim that I am benefited *now* if (1) I want my body to be buried rather than cremated, and (2) I believe that my body *will be* buried rather than cremated. Here is a case in which a person's life is claimed to be going better when no rational person could believe—and this person in fact does not believe—that the relevant desire has been satisfied. While this move may or may not have any theoretical value,[19] it appears to

[19] It does, as Heathwood's example demonstrates, allow for the addition of posthumous benefits and harms, but as we saw earlier the right theory will not

come at much too high a cost, as a return to *Loopy Narcissist* should demonstrate. Let us change Loopy Narcissist's story such that nearly all the desires I had during my life that could have been satisfied during my life (i.e., not future desires) were frustrated and that I believed they were frustrated. However, I believed that *all* my desires about the genuflecting masses after my death were going to be satisfied. If we work the numbers in the theory right (which, given the sheer number of people and days involved, would not be hard), then Heathwood's theory will tell us that Loopy Narcissist had perhaps the best life that has ever been lived, even though all I knew *during my life* was rampant desire frustration. This is not plausible. Accordingly, the correct theory will require that the person believe the desired state of affairs *has* obtained or that it merely *does* obtain.[20]

We now turn to the question of whether the right theory will require something in addition, yet related, to believing that the desired state of affairs either has obtained or does obtain. The salient and obvious fact that needs to be evaluated here is that believing a state of affairs obtains does not entail that the state of affairs actually does obtain. The question is, then, does the right theory require, in addition to belief, that it be *true* that the desired state of affairs either has obtained or does obtain?

---

count these toward personal welfare. I cannot explore all of the other potential benefits and costs associated with this move here.

[20] The second option is meant to deal with states of affairs that do not begin to obtain *at* any time, as in the case, for example, of mathematical truths.

We can begin to evaluate the requirement of true belief in this context by starting with an uncontroversial case. Let us begin with the Cartesian-style assumption that we cannot be wrong about how experiences seem to us. In other words, it is not possible to have *experience* X while at the same time believing that one is not having *experience* X.[21] Now let us suppose that I merely desire the *experience* of X—not that the *state of affairs* of X actually obtains—and that I believe I am having the experience of X. Requiring true belief in this case is unproblematic since this is a case in which belief *does* entail truth.

However, the situation is much more problematic when belief does not entail truth, as in every case in which the desire is not merely for the experience but for the desired state of affairs to obtain outside one's head. Heathwood (17), who as mentioned above also incorporates a belief requirement in his desire-satisfaction theory, thinks truth poses such a problem that he proposes dropping it altogether:

> If the theory still requires the desire really to be satisfied, then the Argument from Remote Desires has not gone away. Suppose Parfit comes to believe and to desire that the stranger has been cured. Then the proposal under consideration will imply that whether Parfit's life is made better depends upon the further issue of whether the stranger has really been cured, a state of affairs remote to Parfit. If one was moved by Parfit's original case, one must be equally moved again. Just tacking a belief requirement onto a traditional desire theory therefore isn't enough to avoid the problem of remote desires after all.

---

[21] Of course, it would be possible to have experience X while at the same time believing that one is *only* experiencing X and that the state of affairs corresponding to experience X is not occurring in reality.

What is not clear in this passage is in what way the stranger's being cured is "a state of affairs remote to Parfit." Is it because Parfit comes to believe that the stranger has been cured on the basis of no evidence whatsoever? If so, then there are avenues left open to a desire-*satisfaction* approach—short of amputating the truth requirement altogether—that merit consideration. I will not pursue any strategy of that sort here since I take Heathwood to mean something much more problematic for a desire-satisfaction theory of personal welfare. In claiming that "a life in the experience machine is just as good for the person who lives it as the corresponding non-hallucinatory life," Heathwood (36) is in essence claiming that all states of affairs beyond our own experiences are too remote from us to impact our well-being.

Thus Heathwood drops the requirement—in a purported desire-*satisfaction* theory—that desires be satisfied in two ways. First, as we saw earlier, it need not even be *possible* that a desire be satisfied as long as the person believes it *will* be satisfied in the future. Second, a desire need never *actually* be satisfied as long as the person believes it either has been or will be satisfied in the future. Accordingly, the theory entails that there is no well-being-related reason not to enter the experience machine (Heathwood: 36 n.52).

Three points about this aspect of Heathwood's theory merit discussion here. First, this theory faces a much more concrete and prevalent problem than evaluating lives lived in experience machines. Dreams seem to present a problem here that would not plague other desire-based theories. Traditional

desire-satisfaction theory could easily handle counting our desires *about* our dreams toward our welfare and safely ignore our desires *in* our dreams as they, *qua* dream desires, are not satisfied or frustrated in reality. However, if there is no welfare-related reason not to enter the experience machine, then it is unclear on what basis Heathwood would exclude dream desire satisfaction and frustration. This will lead to some counter-intuitive results concerning welfare (e.g., the *Loopy Narcissist* will be a great life once again if we merely shift his posthumous desires to his dream life).

Second, there does seem to be a compelling well-being-related reason not to enter the experience machine. The reason is that a great deal of what people actually do *care* about are not experiences at all, but rather are things outside of their own heads. And it is only as a result of caring about something that the life of a person can go better or worse at all. This idea will be fully fleshed out in the next chapter.

Finally, dropping the truth component from desire satisfactionism causes the theory to suffer from a relatively subtle form of paternalism, which does seem to violate our Principle Concerning Paternalism.[22] At first, though, the theory is very liberal; a large subset of one's actual desires counts toward well-being. Moreover, this subset of desires is not limited to one's experiences, as "desires about the external world count, too" (Heathwood 18 n.30). The problem is that, although desires about the external world *do* count, the actual

[22] PCP: Paternalistic claims in axiology must be justified by a compelling theoretical interest and must be narrowly tailored to serve that interest.

external world itself *does not* count. The paternalism comes in the form of reinterpreting our desires for us.[23] If I were to say that I have welfare-related reasons for having my desire for X satisfied, the theory tells me not to be silly—what I *really* have are welfare-related reasons for the experience of having the desire for X satisfied. This response would likely strike those with these kinds of desires as deeply unsatisfying since this is really not what they want at all.

However, what if there were a solution to this issue that did not involve any sort of paternalism (in keeping with the requirements of PCP)? The most obvious solution of this sort is simply to resort to the fact of the matter as to what the person in question actually cares about. In other words, is it enough for you merely to *believe* that this particular desire has been satisfied (i.e., you just want the experience of X), or do you care both to believe the desire has been satisfied *and* to actually have the desired state of affairs obtain (i.e., you want the experience of X and for the state of affairs X to actually obtain)? Another way to understand this question is to see it simply as an effort to find out what the person *really* cares about. For example, some people (likely future philosophers) really care about only the experience of having their

[23] Interestingly, standard forms of desire theory may also be guilty of a similar kind of paternalism by *not* reinterpreting our desires for us. So if I claim that I desire to own Jack Rabbit Slim's and the theory takes my desire at face value, then it will be good for me only if I actually come to own Jack Rabbit Slim's. However, perhaps it is the case that I only want the *experience* of owning Jack Rabbit Slim's even though I claim to want to actually own it. The problem here is that the theory is telling me, paternalistically, that the actual state of affairs affects my welfare when all I really wanted was just the experience.

desires satisfied, but are unaware of this fact because they have never had thought experiments like the experience machine brought to their attention. This emphasis on getting to the bottom of what the person actually cares about yields desires that are much less problematic from a paternalism standpoint. For those desires in which the object is just the experience, the truth requirement is entirely unproblematic (as noted above). For those desires in which the object is a state of affairs outside the head, the truth condition is much more controversial. The bottom line here is that what the person *actually* cares about—the experience or the state of affairs—determines what the object of the desire is for purposes of calculating welfare, thus allowing the theory to hook into reality, or not, based on the actual desire in question.

The advantage of this approach can be seen by examining an issue raised by Feldman. Although Feldman, a committed hedonist, acknowledges that "pleasure taken in things that are true does seem somehow better than equal pleasure taken in things that are false," he is "puzzled by an apparent disanalogy between pleasure and pain here" (Feldman 2004: 111). He is puzzled because he has "no clear intuitions concerning the impact of the falsity of the object" when it comes to pain, which is in conflict with his intuitions concerning false pleasures (Feldman 2004: 111). While I am inclined to agree with Feldman here, as I would have, other things being equal, no compelling basis for preferring a true pain to a false one (or vice versa), my preference does not alter the theory in a way that would make *my* preference *everyone's* preference. The approach of looking to what the person in question *actually*

cares about will allow my theory to handle everyone's preferences seamlessly and nonpaternalistically on this (and every other) issue.

V. Warm Desires

While the objection from remote desires is an invitation to reduce the ambit of desires that are relevant to welfare in one way, another way to reduce the ambit is to limit the theory to *warm* desires—a term apparently coined by David Lewis (1988: 323) to refer to desires for which "you feel enthusiasm, you take pleasure in the prospect of fulfilment." This is the strategy employed by Heathwood (20-22) in response to this distinction drawn by Lewis and others. An evaluation of this strategy should reveal that either this distinction is correct and needs to be included in the right theory, or that the right theory needs to include the guiding intuition that prompted the move to this distinction.

Notably, in drawing this distinction Heathwood offers no analysis of the central concepts. This complicates the evaluation by requiring us to aim at an undefined target. The first step here, then, will be to construct a target as charitably as we can from the tools we have been provided, which include some examples of both warm and cold desires and some quotes from other philosophers who have drawn a similar distinction. Heathwood (20) provides the following examples of warm desires:

- anticipating a job interview tomorrow and strongly wanting it to go well,
- wanting so badly to go back to sleep when one's alarm clock sounds,

- checking the newspaper and hoping intently that one's candidate has won the election,
- dreaming about one day owning one's own business.

Heathwood (20) then gives a list of examples of cold desires:

- preferring, after all, to let one's guest have the last piece of pizza,
- forcing oneself to get out of bed despite how tired one is,
- saying No to a cigarette, despite its appeal, because one is trying to quit,
- deciding to continue slogging through a tedious article.

After giving these examples, Heathwood (20) goes on to cite with approval three philosophers who have drawn "more or less the distinction illustrated by the examples above." Sumner (1996: 121), the only one of the three addressing the issue of welfare, draws a distinction between wanting in the "attitudinal sense" (Heathwood's warm desires) and wanting in the "behavioural sense" (Heathwood's cold desires). "Wanting to do something in the behavioural sense, is just having some reason or other for doing it, with no restriction whatever placed on the range of possible reasons" (Sumner 1996: 121). Examples of behavioural motivations, according to Sumner, are altruism, a sense of obligation, or doing what we feel we ought or must. In the attitudinal sense, on the other hand, "wanting to do something requires finding the prospect of it pleasing or agreeable, or welcoming the opportunity to do it, or looking forward to it with gusto or enthusiasm" (Sumner 1996: 121). In other

words, the attitudinal sense of desire is a subset of the behavioural sense that contains, roughly, just those desires we, in some sense, like.

G.F. Schueler (1995: 35) makes a similar distinction between what he calls "desires proper" (Heathwood's warm desires) and "pro attitudes" (Heathwood's cold desires). Desires proper "will presumably include such things as cravings, urges, wishes, hopes, yens, and the like, as well as at least some motivated desires, but not such things as moral or political beliefs that could appear in practical deliberation as arguing against the dictates of ones [sic] urges, cravings, or wishes" (Schueler 1995: 35). Pro attitude, by contrast, "refers to *whatever* led the agent to perform that action" (Schueler 1995: 35).[24]

---

[24] One problem, which is not particularly relevant for our purposes, is that these descriptions of the two senses of desire do not square with the descriptions Schueler gives on page one of his book, which Heathwood cites approvingly in full:

> The aim of this book is to try to understand how, and indeed whether, desires can have a role in practical reason and the explanation of intentional action. To that end a rather simple and (I think) obvious distinction is explained in chapter 1 and then put to work in the following chapters. The distinction is that between two senses of the term "desire": On the one side is what might be called the philosophers' sense, in which, as G. E. M. Anscombe (1963, 68) says, "the primitive sign of wanting is trying to get," that is, the sense in which desires are so to speak automatically tied to actions because the term "desire" is understood so broadly as to apply to whatever moves someone to act. On the other side is the more ordinary sense, in which one can do things one has no desire to do, that is, the sense in which one can reflect on one's own desires, try to figure out what one wants, compare one's own desires with the desires of others or the requirements of morals, the law, etiquette or prudence, and in the end, perhaps, even decide that some desires one has, even very strong ones, shouldn't be acted on at all. (Schueler 1995: 1)

The problem is that the "more ordinary sense" to which Schueler refers—what he later calls desires proper and what Heathwood calls warm desires—includes

In short, desires proper are a subset of pro attitudes that do include certain types of value judgments.

Finally, Heathwood cites Wayne Davis, who distinguishes between the *volitive* sense of desire (Heathwood's cold desires) and the *appetitive* sense of desire (Heathwood's warm desires). Davis (1984: 181-82) claims that volitive desire is synonymous with "want, wish, and would like," while appetitive desires has "the near synonyms appetite, hungering, craving, yearning, longing, and urge." Moreover, Davis (1984: 186) claims that "volitive desires are typically based on reasons," in this sense resembling beliefs, whereas appetitive desires "are not the sorts of things we have reasons for or against," and in this sense "are more like aches and pains." In addition, although both senses of desire influence action, "volitive desire is a more reliable indicator of action" (Davis 1984: 187). "Appetitive desire, on the other hand, is a more reliable indicator of enjoyment," although the satisfaction of a desire in either sense "tends to be enjoyable" (Davis 1984: 187). Finally, Davis (1984: 183-88) claims that the objects of appetitive desire are "appealing" and "viewed with pleasure," while volitive desires are powerfully influenced by "value-judgments" and are "manifestations of the will."[25]

"morals" and "law," which are specifically excluded in the definition of desires proper given above.

[25] One difference between Davis, on the one hand, and Sumner and Schueler on the other is that Davis claims his two senses of desire are logically independent (i.e., one category is not a subset of the other). However, the cost of this logical independence in terms of violence to the language is high, in that it allows him to claim coherence for such gems as "I desire a hammer, but do

Now that we have presented all the evidence it is time to look for the common thread. What, then, is the essential feature of warm desires that makes them, according to Heathwood's theory, relevant to welfare? The most obvious answer is that the essence of warm desires is that, when satisfied, they increase one's welfare. However, this cannot be the correct analysis, as this would make the appeal to warm desires viciously circular, at least when it comes to their use in a theory of welfare. The actual essence of warm desires is much more subtle, but it becomes clear if one carefully examines the evidence while keeping the general approach of desire theory—satisfied desires increase well-being—firmly in mind. Once this is done, it becomes apparent that Heathwood's warm desires are simply those desires that express what the agent *really wants to do*.[26] And while I think the appeal to warm desires, so defined, fails, a closer look at this strategy will prove fruitful.

The first thing to notice about defining warm desires as those that express what the agent really wants to do is that this definition can be interpreted in different ways. Moreover, this is not merely an academic exercise, as I think Heathwood's examples and the literature he cites employ at least two different senses of this phrase. Let us start with the sense in which I

---

not have a desire for a hammer," and, "We desire to eat, but do not have a desire to eat" (Davis 1984: 184).

[26] Heathwood (24) seems to endorse just such an interpretation in the explanation of a hypothetical that "exploit[s] the notion of warmth of desire," which reads in relevant part: "For it would be very natural to hear me put it like this: 'I don't really *want* to be feeling this sensation, but I have to feel it in order to avoid infection, so give it to me,' while Father would not say anything analogous. He *does* 'really want' to see A's."

understand (or, at least, as I take Heathwood to understand) Sumner, Schueler, and Davis to be using this phrase. These three seem to be using the phrase roughly to convey the idea that this desire reflects what the agent wants to do before the agent takes into account the broader context generally and other people in particular. Both Schueler and Davis use the words *craving* and *urge* (Sumner talks about *relish* and *gusto*) in relation to Heathwood's warm desires. These warm desires become cold once they are introduced to the buzz kills of *ought* and *must* (Sumner), *responsibility* and *moral obligation* (Schueler), and *reasons* and *value judgments* (Davis).[27]

Does this understanding of warm desires succeed in ensuring that the satisfaction of these desires will increase personal welfare? No. Ultimately this attempt fails, as an example from Davis should help illustrate. Using Heathwood's terminology, Davis claims that cold desires are manifestations of the will and that both warm desires and value judgments act on the will. Weakness of will, according to Davis, occurs when warm desires and value judgments come into conflict and the warm desire wins out over the value judgment in motivating the action that is actually taken. Accordingly, under Heathwood's theory, when a person displays weakness of will (leaving aside the issue of any resulting future desire satisfactions and frustrations and given

---

[27] Heathwood (26) seems to endorse this understanding of cold desires in claiming that "what we most prefer in the cold, rationalistic way is heavily influenced by our values—by what we think would be impersonally good, or just, or right, or otherwise worthy of being preferred." As we will see at the end of Chapter Five, the claim that the satisfaction of these types of desires cannot enhance welfare is untenable.

Davis's account of this phenomenon), this will make that person's life go better. It might prove to be an interesting debate as to whether weakness of will *ever* increases personal welfare, but it should be an exceedingly short and uninteresting debate as to whether it *always* does so. It is not plausible to claim that it does, and this helps to underline the reason that the satisfaction of warm desires, understood in this sense, fails to increase welfare. In addition to the problem of weakness of will, it would be an odd result if value judgments generally and moral judgments in particular—the very essence of cold desires—were always irrelevant to determining one's welfare. Many historical figures—Lincoln, Gandhi, MLK—might be rather surprised by such a claim! Equally surprising is the claim that the stuff that bubbles up from our ids—cravings, urges, and the like—serves as the basis for our well-being. If this is what is meant by warm desires—those desires that express what the agent really wants to do—then it is a nonstarter.

There is, however, a more plausible interpretation of warm desires that can be gleaned from Heathwood's examples. What the agent really wants to do in each of his examples has much less to do with values in general and morality in particular. While there is still conflict in the cold desires, it is more value neutral, at least on its face. If we begin by examining the four examples of warm desires, it is easy to see each one as expressing what the agent really wants to do, as there are no readily apparent conflicting desires. The warm desires also seem less problematic because they appear, for the most part, to be the product of reasoned thought as opposed to the id's urges and cravings.

So far so good, but it is not yet clear what work the addition of the warm desires distinction is performing, as these desires would be included in traditional versions of desire theory. The weakness in this approach, however, becomes apparent when we focus on the four examples of cold desires. In each of these cases, there is a very clear conflict between two desires that cannot both be satisfied. The traditional desire theory approach would be to look at the intensity of each desire and then subtract the intensity of the frustrated desire from the intensity of the satisfied desire in order to determine the impact of this episode on the person's welfare. This is not the approach Heathwood takes, as we are told these cold desires *are not relevant* to welfare. What, then, makes these desires cold so that they are completely dismissed from welfare calculations? One possibility is that anytime there is an inherent conflict the desires are dismissed as being cold. This cannot be the case, as desires squarely within any desire theory often come with some degree, however minor, of inherent conflict (e.g., desires concerning personal relationships, careers, etc.). Since these desires are being singled out for exclusion, the traditional approach is being eschewed, and inherent conflict cannot be sufficient to create a cold desire, the only remaining possibility is that—upon satisfaction of a cold desire—the person is not getting what he really wants due to the person's choosing to satisfy the less intense of the two conflicting desires. If this were not the case, how could satisfying the stronger desire be classified as the person's not getting what he really wants? An examination of each of the cases reveals this to be the case. The clear indication is that the weaker desire

in each of these cases is being frustrated at the expense of the stronger. So what the person really wants—to eat the last piece of pizza, to stay in bed, to smoke the cigarette, to quit reading—is not what he ends up getting.

Cold desires, then, are those desires which, when satisfied, do not result in the person getting what he really wants because he chooses to satisfy the weaker of two conflicting desires. This approach obviously raises some questions. First, why *exclude* such desires from welfare calculation? A traditional desire theory would claim the satisfaction of cold desires so understood would make a life go worse. Why is this not the case? Could it be that *every* situation involving cold desires, no matter if they are frustrated or satisfied, can safely be ignored for personal welfare purposes? This is a claim in need of a justification. Second, if this really reflects the situation—if a person really chooses to satisfy a less intense desire—*why* do they do this? This is difficult to explain unless one makes use of a hierarchical model of the will such as Frankfurt's. And it just so happens that by resorting to the tools provided by Frankfurt in the last chapter, we can capture the central intuition behind the appeal to warm desires in a very clear and straightforward manner.

VI. Third Fix: Caring

Chapters One and Two repeatedly made the case for caring being a central component in any theory of welfare. Chapter Three, among other things, provided an in-depth analysis of caring and related concepts. This section will show how the concept of caring can be deployed in a straightforward manner to solve the problems that have traditionally plagued

other desire-satisfaction theories. Specifically, we will see how it helps with issues related to warm desires, ideal desires, and several types of remote desires, as well as some distinctions that have not yet been introduced.

i. Warm Desires

Before we get to these other topics, let us continue our discussion of warm desires in order to see what light caring can shed here. Note that we determined the warm desires of Sumner, Schueler, and Davis were best understood as desires that express what we really want before factoring in value judgments in general and morality in particular.[28] The question, then, is whether or not it is plausible to claim that desires that factor in value judgments (Heathwood's cold desires) are not even relevant to well-being. As was suggested in the last section, this would be a very odd result. It is very common for us to value our significant others, children, parents, siblings, friends, careers, and hobbies. Yet cold desires, so understood, would make our desires concerning all of these things *irrelevant* to our welfare. This view is not plausible and should be rejected, as it is *precisely* because of the things we value, or *care* about, that it is possible for our lives to go better or worse.

---

[28] It could be the case that warm desires are those that express what we really want before factoring in either value judgments or morality, but not both. Such a position is untenable, I think, if we take the (proper) view of morality as expressing one's value judgments about people (i.e., the weight we assign to the interests of others in deciding upon a course of action). Taking this view, it is very hard to see a plausible motivation for excluding value judgments about people from warm desires while including value judgments about nonpersons, and vice versa.

Next, we considered the possibility that cold desires could be defined as those involving incompatible, conflicting desires where the person chooses to satisfy the weaker desire at the expense of frustrating the stronger. Two problems with this approach were immediately identified. First, why is the satisfaction of the cold desires described previously irrelevant to one's welfare, as Heathwood claims, as opposed to being detrimental to one's welfare, as a traditional desire theory would claim? Second, *why* would one choose to satisfy a weaker desire at the expense of a stronger desire? This seems difficult to explain unless an odd definition of desire is being used or relevant details are being omitted, or both. However, both of these questions can be readily answered by deploying the conceptual framework set out by Frankfurt in the last chapter. Here is how Frankfurt would explain Heathwood's (20) cold desire—"saying No to a cigarette, despite its appeal, because one is trying to quit"—described above. The person, let's call him Hitchens, has two incompatible first-order desires—a desire not to smoke and a stronger desire to smoke. In addition, due to a recent health scare, Hitchens has a strong and unconflicted second-order volition that his first-order desire not to smoke be effective, thus constituting his will. Now we have an answer for both of our questions. First, Hitchens does not smoke because he wants to want not to smoke, despite his stronger first-order desire to smoke. Second, this state of affairs does not make Hitchens's life go worse (as a traditional desire theory might claim), nor is it irrelevant to his welfare (as Heathwood's theory does claim). Instead, assuming that Hitchens *cares* about his desire not to smoke,

this sequence makes his life go better, as he is getting what he cares about, he has the will he wants to have, he is exercising his free will, etc. Accordingly, in the context of the welfare debate, Frankfurt's concept of caring both embraces the intuition behind the warm/cold desire distinction and provides a better explanation for the entire array of warm and cold desires, no matter what the proper analysis of those terms turns out to be.

ii. Ideal Desires

Caring also resolves a lingering issue raised during the discussion of idealized desires. Recall that the appeal to idealized desires was in response to the problem of defective desires. Assuming the move from actual desires to idealized desires is not a mere Trojan Horse for objective-list theories, then how can desires be "defective"? Since it cannot mean that some objects cannot be desired, we are left with two remaining possibilities. The first of these we examined in the section on future desires, where we concluded that a desire could be defective because, in satisfying the desire, we will be frustrating more of our actual future desires than if we had not satisfied the desire. We could think of this as a situation in which the object of our desire is not something one "really" wants, in that it will make our lives go worse by frustrating our actual future desires. The second way in which a desire may be characterized as being defective could also be characterized as a case in which the object of our desire is not something we really want. Although this possibility was mentioned in the section called "Caring" in the last chapter, a slight return adapted for this particular context will be instructive. Our desires or, more

specifically, the things we care about, may be defective in that the objects are not things we really want, or care about, because we either do not know ourselves well enough, do not know the objects well enough, or both.

Not knowing the objects of our desire well enough should not be a surprising phenomenon. After all, some of the things we want are things we have never had, so it should not be surprising to learn that the object of our desire was different from what we imagined it to be. This may occur in one of two ways. First, we may notice that one of the things we care about conflicts with another one. Frankfurt (2006: 49) gives the example of a person who cares about worldly success and about peace of mind, only to discover that pursuing one tends to interfere with attaining the other: "As we learn more about what each is and what it entails, it will often become clear that one arouses in us a more substantial interest and concern than the other." Second, we may discover that we just do not understand the thing we want well enough, independent of any conflict in the things we want. We may discover either that the object is just different than we thought it was or that the consequences and requirements of caring about it differ from what we had supposed (Frankfurt 2006: 49).

Not knowing ourselves very well should not be that surprising either. Frankfurt puts it quite well:

> Our motives and our dispositions are notoriously uncertain and opaque, and we often get ourselves wrong. It is hard to be sure what we can bring ourselves to do, or how we will behave when the chips are down. The will is a thing as real as any reality outside us. The truth about it does not depend upon what we think it is, or upon what we wish it were. (Frankfurt 2006: 49-50)

The strategy, then, for dealing with desires that are defective (i.e., those desires that do not express what we really care about) because we do not understand some combination of the object or ourselves very well will be as follows. Following Heathwood (25):

> I will assume that desires and beliefs are very brief entities – so brief that there is no time for their intensity to change. What we would normally describe as a change in intensity of a single desire or belief we will, for the purposes of the theory, describe as an occurrence of a new desire or belief of a different intensity for the same [state of affairs].

Defined in this way, the theory (which will be explained in much greater detail in the next chapter) will reach the right conclusion, in terms of impact on welfare, for the entire range of cases in which what we care about changes over time. The most obvious case is when we get either ourselves or the object wrong such that when we get the object, we no longer want it. The theory will claim that the impact on our welfare decreases exactly as fast as we determine that the object is not something we care about after all. The theory will reach the same result in cases in which what we care about changes over a much longer time period. For example, suppose that over a 25-year period you cease caring about your significant other, whom you cared about very deeply at one time. The theory will claim that the impact on your personal welfare of having that person in your life exactly tracked your decreasing level of caring about that person during those 25 years. Thus, the problem that idealized desires handled very poorly, the concept of caring can solve easily.

iii. Remote Desires

Next, caring resolves the issues left over from the discussion of remote desires. Earlier in this chapter, we discussed a resolution to one type of remote desires; namely, that some desired states of affairs are remote from what we are aware of. The resolution was that the desirer had to have a true belief about whether the desire was actually satisfied. While this is a solution for desires that are remote from our experience, it will not solve the other types of remote desires that were mentioned. The first included desires that seemed to be remote from what a person could care about, as in the case of Kagan's wanting the number of atoms in the universe to be a prime number. Although it should be fairly easy to see how caring will resolve issues dealing with desires that are remote from what we care about, a closer look at this case is in order. First, it seems to combine two types of remote desires in one desire. In other words, not only does it seem remote from what Kagan would care about, but also it *is* remote from what Kagan could realistically expect to be aware of. Although the combination of true belief and caring is sufficient to deal with this case effectively, separating these two issues for the purposes of this topic will help to simplify the analysis. Accordingly, let us change the example to eliminate the experience-requirement issue while keeping the difficulty-caring-about-prime-numbers issue. Kagan, in the new scenario, is still a huge fan of prime numbers, but he is also a huge fan of the St. Louis Cardinals. What Kagan wants is for the Cards' total number of runs scored in the 2011 regular

season to be a prime number. When Albert Pujols[29] scores the 1,229th run, a prime number and a new MLB record, with a walk-off homerun in the bottom of the ninth inning in Houston on September 28th (the last game of the season), Kagan is simply elated.

Now those who would deny that this desire satisfaction increased Kagan's welfare have two possible options left at this point in the proceedings. First, the opponent could simply deny that satisfactions such as this one can increase welfare. But what feature of this desire could one single out as making it irrelevant to welfare? The fact that it involves a number? A prime number? A sports team? If all these possible exceptions seem *ad hoc*, that is because they would be, as our next option should help reinforce. The second option for the opponent is not to deny that this desire could impact one's welfare if someone actually cared about it, but to deny that it would ever actually impact anyone's welfare because, in fact, no one could care about it. When stated so clearly, this empirical claim does not look plausible because of its absolute nature (i.e., *no one* could *ever* care about this). However, I think this is precisely where Kagan's hypothetical gets the bulk of its intuitive force. We just cannot get to a place, mentally, where we can imagine caring about such things. To those people I would simply ask them to feed their imaginations a bit. If it is actually the case that you do not care about anything that would strike a cross-section of your acquaintances as odd, then spend an afternoon at the library picking up random books to get a sense of

[29] Fact: Pujols is the best player in MLB history through the first ten seasons of a career.

what some other people actually do care about.[30] Or peruse random videos on YouTube to get a sense of what people care about. I promise you there is a very large number of people who care deeply about something that you have never heard of—and some of these things will most likely strike you as being at least as hard to care about as the number of runs a baseball team scores in a season. It is just a fact about humans that we care about all sorts of stuff. If it is a thing humans are aware of, then it is likely a thing someone cares about. Sectioning off some subset of these things as being necessarily irrelevant to welfare will either involve some highly implausible metaphysical claims (e.g., objective-list theories) or some highly implausible empirical claims (e.g., no one could ever care about X).

Caring also provides a solution for all the remaining types of remote desires discussed previously. While these desires are sometimes described, as they were earlier in this chapter, as being too remote in *time* to impact welfare, a more accurate description is that these desires are just additional types of desires that are remote from what we care about. Recall that the first type of desire described as being remote in time included my desire, as a seven-year-old kid, to have Davey the gorilla—on orders from The Talking Walnut—beat up the skateboard bullies who were harassing me. The problem is that, by the time my desire is satisfied as a Festivus miracle 75 years later, I have long forgotten about this desire and now feel sorry for the bloodied senior citizens

[30] The writer obviously cared about the topic, as she wrote an entire book on the subject. Moreover, the library carries the book, so its staff likely thinks that other people care about it as well.

who have long since traded in their skateboards for wheelchairs. However, according to standard desire theory, my life is going better now because (1) I had this desire, and (2) the desire was satisfied. This is not plausible. It seems dogmatic to claim that a desire I had when I was seven, and subsequently forgot about, could make my life go better if satisfied when I am 82.

We could characterize this situation as representing a problem with changing desires, as Brandt (1982: 179) does when he writes:

> The fundamental difficulty for the desire-satisfaction theory is that desires change over time: Some occurrence I now want to have happen may be something I did not want to have happen in the past, and will wish had not happened, if it does happen, in the future.

However, the concept of changing desires, without more, does not present any particular difficulty for a properly constructed desire-satisfaction theory. To see this, consider a slight modification to the case of Davey and The Talking Walnut. Keeping all other facts the same, now suppose I had the desire for Davey-delivered retribution for the whole year I was seven, and then, after the skateboard bullies became my friends, I had an equally strong desire for the whole year I was nine that no harm from Davey would befall them. Both desires lasted for exactly one year, and both have long been forgotten when Davey visits my former enemies/friends at the Happy, Happy, Joy, Joy Retirement Center. Since my desire changed before it was satisfied, does this complicate our analysis of the impact of the beatings on my welfare when I am 82? It should not. If the forgotten pro-beating desire is not relevant, it is hard to see why my forgotten anti-beating desire could be relevant. Nor does it seem

to make any difference that my desire changed to one that directly opposed the first desire. For purposes of this analysis, the only relevant detail seems to be that I had two long-forgotten desires. Indeed, it does not even appear to be relevant that one was subsequently satisfied and the other frustrated. This analysis of the issue should become clearer if we think back to the problem with idealized desire-satisfaction theories; specifically, their claim that a person's welfare can be increased not only by getting something that he does not care about, but by getting something that he does not even want. Here, if we count these past desires toward welfare, we would be doing the same thing—giving the person something he does not want and counting it as enhancing his well-being. And it does not appear to matter whether the person remembers having the desire or not because, either way, it is not something he currently wants or cares about. Should it matter whether he merely *remembers* having had a desire for the object at some point in the past?

> One solution to this issue, suggested by Parfit (1984: 151), is as follows:
>
> > Some desires are implicitly conditional on their own persistence. If I now want to swim when the Moon later rises, I may want to do so only if, when the Moon rises, I still want to swim. If a desire is conditional on its own persistence, it can obviously be ignored once it is past.

This approach suggests dividing desires into two camps—those that are conditional on their own persistence and those that are not. The former would be counted toward welfare only if they still persist at the time they are satisfied,

while the latter would be counted toward welfare regardless of whether the person still has the desire when satisfied.[31]

This proposal fails for a number of reasons. The first comes to light if we examine the basic idea behind desire theory. The most plausible basic form of desire theory claims, roughly, that a person is benefited when what he desires obtains *while he still wants it*. We can call this a *concurrence requirement* whereby there is a temporal overlap between the desire and desired state of affairs.[32] Desires that are not conditional on their own persistence will struggle with this requirement, as the options here should make clear. The first option is just to claim that there is no necessary overlap between the desire and the desired state of affairs. The problem here is that it is not clear how the occurrence of a state of affairs benefits the person, who, by hypothesis, no longer wants the state of affairs in question. The second option is to claim that there is indeed an overlap of some sort between the desire and the desired state of affairs. However, it seems as though this sort of claim will require some interesting metaphysics. There is no mystery concerning the metaphysical status of desires that are conditional on their own persistence. These are desires that a person has and then, at some point (at least at death if not before), no longer has. So these desires exist in the mind for a time and then

[31] For both classes of desires, any other requirements set out in the theory would have to be satisfied before the desire in question would count toward welfare.

[32] The justification for this requirement stems from the problems, detailed below, associated with the claim that giving a person something that she does not currently want is good for the person.

exist merely as historical facts. The same simple story does not apply for desires that are *not* conditional on their own persistence. Here there are serious questions about both *how* one is created and *what* it is.[33]

Now let us suppose that we figure out how to get one of these things created. We have a desire that is not conditional on its own persistence. *What* is it after the desirer no longer has the desire? It is a desire that is not currently had by the desirer, yet it is still relevant until such a time that it is either satisfied or frustrated. The idea seems to be that these are desires, not located entirely in the past or in any present mind, that still possess the welfare *oomph* (technical term) of a live desire. It is almost as if they are supposed to be a chunk of a soul or a fragment of a ghost. Accordingly, it is easy to see the way in which the *how* of these things being created is complicated by the *what* of their actual existence. I had wanted to avoid the quip that perhaps one needed to utter some magic words to bring one into existence, but perhaps magic words are required when conjuring up magical entities. The bottom line here is that this picture of how these desires meet the concurrence requirement is a bit too fanciful to believe.

---

[33] Parfit sheds precious little light on either question. As for how one of these desires is created, he cites nothing and gives only two examples of desires that are not conditional on their own persistence. The only common thread between Parfit's two examples is that the desirer will never *know* if the desire has been satisfied or frustrated (in one case the person to whom the desire relates is one the desirer will never meet again, and in the other case the desirer will be dead when the desire is either frustrated or satisfied). Since there appears to be no reason at all to suppose that all desires that are not conditional on their own persistence involve cases where the desirer will never know if the desire was ultimately satisfied or frustrated, these examples are of little help.

While I admit the welfare justification for, and metaphysics of, desires that are not conditional on their own persistence makes my brain hurt, perhaps I am just not clever enough to get it. That is fine, because at least three problems remain. The first is an issue we have seen before, and it stems from the fact that these desires do not operate in the way other desires do. We are accustomed to being able to turn a desire OFF; we have a desire—it comes ON—and then we either forget the desire or change our minds—-the desire goes OFF. However, what we have in this case is a desire, by definition, permanently stuck in the ON position.[34] Therefore, when (not if) the person no longer has this desire, the solution proposed would entail that we make the person's life go better by giving her what she no longer wants. The fact that this comes from a prior version of herself—as opposed to either an objective list or an idealized version of herself—does not matter. We do her no favors by giving her what she does not either want or care about.[35]

Another problem stems from the fact that the desirer need not know of the satisfaction or frustration of the desire that is not conditional on its own persistence. As we saw earlier in this chapter, it is hard to see how things we are not aware of can, in and of themselves, make our lives go better for us. Nor will it help here to add a belief or knowledge requirement. Suppose I conjure

---

[34] I say *permanently* stuck in the ON position because, once again, it is very unclear how the eventual satisfaction or frustration of these desires would alter their metaphysical states in such a way as to either destroy them or turn them off. My brain might literally explode if I ever heard someone try to explain this.

[35] Heathwood (12) dismisses the appeal to these desires for the same reason.

up one of these desires, forget about it, and then years later when the desired state of affairs comes to be, I learn of this event. On what basis will we claim that my life is going better as result of my learning about some state of affairs that I care nothing about?

This issue concerning knowledge or belief brings up a third problem, which was discussed in the section on remote desires—the problem of posthumous desires as being another example of those desires that are too remote in time to affect personal welfare. Recall the example of my being a Loopy Narcissist with all sorts of desires about the genuflecting masses after my death. Several issues were discussed, but the primary one was the problem of claiming that this was a very good life for *me* if during my life all my desires were frustrated, only to be offset many times over after my death by the satisfaction of my numerous desires by the adoring throngs. However, the resort to desires that are not conditional on their own persistence entails counting posthumous desire satisfactions and frustrations toward well-being. We have desires, no longer residing in the mind of the desirer, that are stuck in the ON position. And since they are not conditional on their own persistence, it does not seem as though any act I could perform would destroy or nullify them. If that is so—which by definition it seems clearly to be—then there looks to be no reason or mechanism that would accomplish this feat upon my death. Thus, the resort to these desires as a solution to desires that are remote in time fails for several reasons.

The solution, once again, is caring. The fundamental problem with past desires, changing desires, and posthumous desires is not that they are too remote in time, but that they are too remote from what we *care about*. In the case of Davey and The Talking Walnut, the problem is not with time, but rather with the fact that this is no longer something I care about. If I did care about it for the intervening 75 years, then of course it is reasonable to claim that my relishing Davey's brutality makes my life go better. The same applies to the Loopy Narcissist example. The problem is not that time has passed, the problem is that I no longer care about anything because I am dead. If I were still alive and still cared about such things, then it does make my life go better to be worshipped.

iv. Occurrent Desires

Here we can highlight another advantage of limiting a desire theory to those desires that we care about. Limiting the theory in this way ensures that it will only include what Heathwood (19) calls *occurrent* desires, which are desires "the object of which is currently, in some sense, 'before the subject's mind.'" Heathwood contrasts occurrent desires with *dispositional* desires—desires that are "in some sense in the desirer's unconscious mind, and which would, at least in normal cases, become occurrent if the subject were to think about the proposition in question"—and claims that the correct theory will only count occurrent desires toward welfare. Moreover, Heathwood (20) notes that since the correct theory also makes use of the concept of belief, that, too, must be taken in its occurrent sense:

> For there are indefinitely many dispositional desire/ dispositional belief pairs that coincide in us at any moment. For example, each of us, at every moment of our adult lives, dispositionally desires and believes that she will not be killed in one minute by a falling meteor. But it seems to me absurd to suppose our lives are continually made better by this fact – not to mention the millions of others like it of which we, and no one, will ever become aware. In fact, I see no reason to think there are not *infinitely many* propositions dispositionally desired and believed by each of us at each moment of our lives. If we allow them all to count, we may have a hard time explaining how any two actual lives could differ in well-being.

I agree with Heathwood's conclusion and reasoning on this point. However, the advantage of resorting to the concept of caring rather than the concept of occurrent desires becomes clear if we follow Heathwood's (22) meteor example a bit further:

> That earlier question was, Did I just make your life better by causing you to desire and believe occurrently and simultaneously that you will not be killed in one minute by a falling meteor? Probably not, or at least not very much. This is because the desire in question probably wasn't a warm desire, or if there was a warm desire on the scene, it was probably very weak. Interestingly, if you received convincing evidence that a giant meteor was heading for your block, and truly came to believe it and appreciate it, your desire that you not soon be killed by a falling meteor would (if you are like me, at least) become very warm indeed. You'd be frantic, and frantically trying to flee. Such a state is an intrinsically bad state to be in, both intuitively and according to [Heathwood's theory], and the reason, according to [Heathwood's theory], is that the state is a subjective frustration of a warm desire. If you later came to learn that there is no meteor after all, your warm desire that there be no meteor would remain, but the associated belief would change: it would then align with the desire (you'd be wanting and believing the same thing). Such relief is an intrinsically good state to be in, both intuitively and according to [Heathwood's theory] – it is a subjective satisfaction of a warm desire.

Here we see that Heathwood's theory has to appeal to both the concept of occurrent desire and the concept of warm desire (a concept that, as we have

seen, is not clear or successful) in order to accomplish what the concept of caring can handle in a simpler and more successful fashion. You are unlikely to care about the possibility of being hit by a meteor when it is no more than, as far as you are concerned, a mere logical possibility. However, when it is elevated to a metaphysical certainty, you are likely to care about this state of affairs quite a bit.

v. Intrinsic Desires

Finally, caring brings some clarity to the debate over whether desire theory should include both intrinsic and extrinsic desires or just intrinsic desires.[36] The intuition behind limiting a theory to include intrinsic desires (i.e., desires for something as an end in itself) and exclude extrinsic desires (i.e., desires for something merely as a means to some other end) is motivated by cases like the following. Suppose I want to watch *The Wire* on DVD. In order to satisfy my intrinsic desire, however, I must form and satisfy several extrinsic desires—to turn on the TV, to turn on the DVD player, to put the DVD in the DVD player, and so on. While it seems plausible to claim that my life is going better for me by getting my (intrinsic) desire to watch *The Wire* satisfied, it does seem a bit strained, at best, to suppose the same goes for my having satisfied my (extrinsic) desire to turn the input setting on my receiver from Cable to DVD.

---

[36] For a discussion of this topic, see Parfit (1984: 117), Brandt (1979: 111), Carson (2000: 155-59), and Heathwood (22-24).

However, the distinction between intrinsic and extrinsic desires is not always so neat. Consider the process I am engaged in at this moment. What I want, as an end in itself, is for my dissertation to be finished. The means to that end is my continuing to sit at this desk, researching various issues related to personal welfare, pondering solutions as I gaze out the window, writing page after brilliant page, etc. The problem with categorizing my desire to have a finished dissertation as wholly intrinsic and my desire to engage in the range of activities necessary to write it as wholly extrinsic is that neither of these descriptions is accurate. First, my desire to have a complete dissertation is also *partly* extrinsic—it is a means to my receiving a Ph.D. as well as being something I desire for its own sake. Second, my desire to write my dissertation is also partly intrinsic. If I awake tomorrow to find my dissertation finished in the exact way I intended and sitting on my desk, I will not be satisfied with this state of affairs because *I* wanted to undertake the activities necessary to produce my dissertation. In other words, I consider at least some of the steps in the process to be ends and not just mere means. However, the same cannot be said about my extrinsic desires relating to watching *The Wire*. If, after I form the desire to watch *The Wire* and before I set about satisfying this desire, I suddenly find myself on the couch watching *The Wire*, I shall have no complaints whatsoever about not getting to bring my purely extrinsic desires to fruition by my own hands (although I will have some nagging metaphysical questions to attend to).

Therefore, Brandt (1979: 111) was right when he suggested limiting desire theory to include those desires that are "at least partly intrinsic." Accordingly, we could add such a requirement to the theory, but a closer examination shows that this is not necessary. We already know that my desires relating to my dissertation are partly intrinsic while those relating to watching *The Wire* are not, but what makes that the case? The difference is that I *care* about writing my dissertation, while I could not care in the least about how a show ends up on my TV. Thus the concept of caring simply and straightforwardly resolves yet another issue in the quest for the correct theory of personal welfare.

VII. No Fix: The South Park Desires

In this chapter we have encountered many different types of desires that are thought to pose problems for desire-satisfaction theories of welfare. Some of these desires are in fact problematic. As a result, several "fixes" for these problems have been advanced. The right theory will have to take into account our future actual desires and not just some form, idealized or otherwise, of our present desires. It will have to count only those satisfactions and frustrations that we believe, truthfully, have taken place instead of counting all frustrations and satisfactions no matter how remote they may be from our experience. Finally, the right theory will include only those desires we care about rather than counting every desire we find ourselves having.

However, there is another category of supposedly defective desires—let us call them the South Park desires—for which there is no fix. Some philosophers

have claimed that these desires cannot be the basis for increasing one's personal welfare. Although the claim is not entirely clear, it seems to be that the satisfaction of these desires actually makes your life go worse, not that these desires are simply not relevant to welfare.[37] I call these the South Park desires because I think it is a near certainty that every sort of desire any philosopher has ever claimed as being defective in this way has been portrayed on *South Park*. In what follows, I will set out some of these desires, but this list is not exhaustive, certainly in terms of examples and probably not in terms of categories either.

In the episode "Scott Tenorman Must Die," Cartman is repeatedly humiliated by Scott Tenorman, causing Cartman to formulate an elaborate strategy of revenge ending with Scott unknowingly eating bits of his own recently murdered parents (also arranged by Cartman) in a bowl of chili. It is claimed that *malicious* desires,[38] such as these, cannot serve to enhance welfare, although Cartman seems quite pleased licking what he calls the "tears of unfathomable sadness" off Scott's face as the episode ends. In the episode "Chickenlover," the Book Mobile driver has sex with several of the town's chickens, although he later claims that this was just an elaborate plot to

---

[37] For an excellent discussion of all these types of desires, see Heathwood (2005: 487-88).

[38] See, e.g., Feldman's (2002: 617) terrorist example.

encourage Officer Barbrady to learn to read.[39] One might claim that *base* or *degrading* desires such as these also cannot make one's life go better.[40] In the episode "Make Love, Not Warcraft," Stan, Kyle, Kenny, and Cartman play *World of Warcraft* for 21 hours a day for two months, all in an attempt to exact personal revenge on a character in the game played by Jenkins, who has been playing the game nearly every hour of every day for the year-and-a-half since the game was released. One might claim that *pointless* desires, such as this one, cannot enhance well-being.[41] Finally, in the episode "Poor and Stupid," Cartman wants to become a NASCAR driver, but is concerned that he is not poor and stupid enough to fulfill his dream. One might want to claim that *tasteless* or *poorly cultivated* desires, such as the desire to drive in circles for a living (or, even worse, *watch* those people who drive in circles for a living), cannot be the stuff of a good life for the one who lives it.[42] And this barely scratches the surface of the desires portrayed on South Park that would keep those inclined to make lists of defective desires busy for quite some time.

---

[39] I think it is a safe bet to stipulate that the desire for chicken sex was at least partly intrinsic due to the fact that a much less extreme tactic could have been used to reach the same goal.

[40] G. E. Moore (1903: § 56) discusses a perpetual indulger in bestiality as an objection to a hedonistic theory of value, but the point is easily adapted to this context.

[41] See, e.g., Rawls's (1999: 379-80) example of a person who wants to count blades of grass instead of study math.

[42] See, e.g., Heathwood's (2005: 488) example of a person's preferring Muzak to Mozart.

In the face of objections based on desires such as these, the desire theorist has two options. He can either admit the theory as it stands is wrong and must be rejected or revised, or he can claim that these desires can serve as the basis for personal welfare. An examination of these options should point us in the right direction. If, as we established long ago, desire theory is the right general approach to welfare, then carving out an exception for these types of desires—in addition to the ones set out already earlier in this chapter—does not look promising. The problem is one of metaphysics. While one could certainly *claim* that these desires cannot enhance well-being, finding *an adequate metaphysical basis* for such a claim is highly doubtful. What would be required is a list of desires (or, perhaps, a list of essential characteristics of desires) that cannot increase personal welfare even though the person *cares* about the desire, *correctly believes* it has been satisfied, and does not cause the frustration of relevant *future* desires. The inherently odd nature of such a list might only be exceeded by the theory of epistemology required to discover it or the story of how it came to be. Therefore, unless one wants to dabble in bold and adventurous metaphysics and epistemology, there is no basis to exclude desires that we may not understand or like, or both.

Frankfurt (2006: 50), in a particularly eloquent passage, sums up this idea nicely:

> Once we have learned as much as possible about the natural characteristics of the things we care about, and as much as possible about ourselves, there are no further substantive corrections that can be made. There is really nothing else to look

> for so far as the normativity of final ends is concerned. There is nothing else to get right.
>
> The legitimacy and the worthiness of our final ends are not susceptible to being demonstrated by impersonal considerations that all rational agents would accept as appropriately controlling. Sometimes, normative disagreements cannot be rationally resolved. It may even be true that other people are required by what they care about to harm or to destroy what we love. Our love may be inspired by an endearing vision of how relationships between individuals might ideally be arranged; but other people may be driven by what they care about to struggle against arranging things in that way. There may be no convincing basis for regarding either them or ourselves as rationally defective or as having made some sort of mistake.
>
> So far as reason goes, the conflict between us may be irreducible. There may be no way to deal with it, in the end, other than to separate or to slug it out. This is a discouraging outcome, but it does not imply a deficiency in my theory. It is just a fact of life.

The bottom line is that some people care about things that we do not. When we appeal to a fact about value to persuade them that they want the "wrong" things, they can simply and conceivably reply that they do not value what we value—no matter what the object is. So what we value helps our lives go better, and what they value helps make their lives go better—end of story (at least as far as personal welfare is concerned). Now it may be that there are other ways to assess a person's life besides welfare. We may be able to assess lives in terms of morality, achievement, excellence, virtue, aesthetic value, dignity, or how good for the world (as opposed to how good for its subject) a life was. If so, then it would make sense to rank the lives of those with South Park desires low on one or more of these scales while still admitting they were good

lives *for the people who lived them*.[43] It would be nice to live in a world where instant karma was a part of the fabric of reality; a place where bad deeds were always and everywhere met with bad outcomes for the perpetrators. I even have an occasional student who really feels that somehow, some way, this must be so. Sadly, this just does not mesh with the facts on the ground—many a tyrannical dictator lives in the lap of luxury to a ripe old age, and many a mensch suffers some unspeakable and untimely end.

In the final chapter, we will (finally!) see my theory in detail, see it put to use, and examine some of its more interesting implications.

[43] This is a common strategy in cases like this. See, e.g., Griffin (1986: 23), Sumner (1996: 20-29), Feldman (2004: 8-12), and Heathwood (2005: 500).

CHAPTER FIVE: CARING SATISFACTIONISM & THE PARADOX OF WELFARE

Some lives go better than other lives for the people who live them. Hopefully, both the truth and the import of this claim were established many pages ago. Further, let us suppose that we can, in theory at least, figure out what makes one life go better than another in order to determine, as between two lives, which one went (or is going or will go) better for the person who lived (or is living or will live) it. Now, if it is true that some lives go better than others and that we can figure this out as described, then the conjunction of these two facts entails some requirements for a theory of personal welfare that should be made explicit.

I. CARING SATISFACTIONISM

First, a theory of personal welfare must *identify* the fundamental bearers of intrinsic value for a subject. In other words, for a theory to count as a theory of welfare at all, it must pick out something that is good in itself for a person to get. This has been the goal of the project to this point. We have examined several possibilities, all with an eye toward finding the thing or things that will serve as the basis of our theory. What we came up with at the end of the last chapter was, roughly, the best life for the person who lives it is the life that best satisfies the desires the person cares about, where the person is also aware of those desire satisfactions.

Second, a complete theory of personal welfare must tell us how to determine the *value* of the fundamental bearers of intrinsic value for a subject.

For if we are to determine which of two lives went better for the subject, then it will not be enough simply to identify the stuff of good lives; we must also be able to figure out exactly how good any particular instance of the stuff is in order to compare lives.

Finally, a theory of welfare should specify how to *calculate* the overall total of intrinsic value for the subject's entire life. Presumably, this will simply consist of summing up the values mentioned in the last paragraph. However, this need not be the case.[1] Perhaps the right theory involves derivatives, exponents, and differential equations, but such an exotic approach would require a very interesting explanation indeed.

Keeping these requirements in mind, we can now state the theory, Caring Satisfactionism ("CS"), in full:

CS(i): Every satisfaction of a caring desire that the desirer believes has been satisfied is intrinsically good for its subject; every frustration of a caring desire that the desirer believes has been frustrated is intrinsically bad for its subject.

CS(ii): The intrinsic value for its subject of a caring desire satisfaction or frustration is a function of the intensities of the component desires and belief.

CS(iii): The intrinsic value of a life (or part of a life) for the one who lives it = the sum of the caring events contained therein.

The first thing to notice about CS(i), CS(ii), and CS(iii) is that they each satisfy one of three requirements for a complete theory of personal welfare

---

[1] J. David Velleman (1991) actually does argue that this is not the case by suggesting that factors like the timing of the basic goods within a life (i.e., the timing of the goods in the overall narrative structure of the life) can affect the value of those goods. I find this claim to be either implausible or one that CS can easily accommodate. I will have to leave this claim undefended here.

enumerated above. CS(i) tells us what is good in itself for a person to get, CS(ii) tells us how good these fundamental bearers of intrinsic value are for a person, and CS(iii) tells us how the values of fundamental bearers go together in assessing the intrinsic value of all or part of a life.

Before turning to an examination of some of the noteworthy features of CS, a thorough explanation of the theory—particularly CS(i) and CS(ii)—is in order. CS(i) requires that two types of things be present for something to be either intrinsically good or intrinsically bad for a subject: One of those requirements, belief, is self-explanatory,[2] while the other, caring desire satisfaction or frustration, needs some additional explanation.

Caring is described at length in Chapter Three. The most pertinent part for our current purpose reads as follows:

> Caring requires more than just having a desire and more than accepting, approving of, or endorsing a desire. Caring requires wanting the desire *sustained*. . . . This focus and attention on the desire owe to the fact that this desire is one with which the person *identifies* himself, and which he accepts as expressing what he really wants. . . . Caring consists in having a higher-order volition [or desire] that the person also wants sustained. (Chapter Three: § VIII. Caring)

Therefore, in order to have a caring desire satisfied, there must be a first-order desire satisfaction that the subject cares about. This can be a desire

---

[2] For purposes of CS, I think even accidental belief is sufficient. While accidental belief—as in the Gettier-type cases—is certainly troubling from an epistemological perspective, there do not appear to be any corresponding concerns here. If the person believes X has obtained and X has, in fact obtained, then this seems to sufficient from a well-being perspective even if the person comes to believe X in some epistemically flawed manner. However, CS could easily be adapted to incorporate a *knowledge* or *awareness* component in place of belief.

satisfaction in either the traditional sense (i.e., the desired state of affairs obtains) or the broader, but still technically accurate, sense described in Chapter Four, where what the subject really cares about is merely the *experience* of the desired state of affairs obtaining, and the subject has the desired experience. CS(i), then, simply claims that when a desire that the subject cares about is satisfied and the subject believes this to be the case, then this is intrinsically good for the subject. In addition, the frustration of a desire the subject cares about, coupled with the corresponding belief, is intrinsically bad for the subject. These satisfactions and frustrations are the "caring events" referenced in CS(iii).

The caring events discussed in CS(i) each contain two desires (a first-order desire and a higher-order desire) and an associated belief. Desires and beliefs vary in intensity from not wanting or believing something all the way to wanting or believing something as much as is possible. CS(ii) simply claims that exactly how good (or bad) the events described in CS(i) are for a person is a function of the intensities of the relevant desires and beliefs.

Now comes the task of specifying exactly how to quantify the theory. As mentioned earlier, this is required in order to allow for a definitive answer to the question of which of two lives went better for the people who lived them. This is also the point at which many theorists stop, and at which many readers will balk at the attempt and begin to quibble with the way in which the theory is quantified. And while a healthy dose of skepticism is perhaps warranted here, it should not cloud what has been accomplished to this point if this is

where one finds fault with the theory. The accomplishment would be the fairly specific outlines of the correct theory of personal welfare, which specifies both what the fundamental bearers of intrinsic value are and generally how good they are for a person.

The first step in the quantification process will be to follow Heathwood (25) in the assumption

> that desires and beliefs are very brief entities—so brief that there is no time for their intensity to change. What we would normally describe as a change in intensity of a single desire or belief we will, for the purposes of the theory, describe as an occurrence of a new desire or belief of a different intensity.

This assumption comes at no theoretical cost and will prevent the need for some difficult math.

The next step in the process of quantifying CS(ii) is to take stock of component parts we have at our disposal. These separate entities—two desires and a belief, each with its own intensity—will comprise each instance of an intrinsic good. Now, for theoretical purposes, let us assume that desires and beliefs can range in intensity from zero (no belief or desire at all) to ten (maximum possible belief or desire) and may fall at any point along that continuum.[3] With these assumptions in hand, the math for the theory becomes fairly straightforward. As argued elsewhere in this project, it is hard to see how events a person does not care about, want, or believe have occurred

[3] For any being that is a person in the Frankfurtian sense, there would seem to be an upper limit on the intensity of desires. Thus, the thought that for a desire of any intensity we could always imagine a more intense desire does not present any difficulties for CS. Alternatively, CS could easily be changed to accommodate the idea that there is no upper limit on the intensity of desires.

can, in and of themselves, make a life go better or worse. And each of these outcomes would be represented by an intensity of zero. Accordingly, if each of these three elements must be present (i.e., must have an intensity above zero), then the most obvious solution is to multiply each of the three numbers together to arrive at the value for CS(ii). This approach will ensure that the utter lack of any of the three will result in no effect upon personal welfare. CS(ii), then, will be determined in any particular instance by multiplying the intensity of the relevant first-order desire by the intensity of the relevant higher-order (caring) desire, and by the intensity of the relevant belief. The product of each instance of a CS(ii) satisfaction will be left positive and the product of each instance of a CS(ii) frustration will be made negative.

The work left for CS(iii) is also fairly straightforward. All that is left for CS(iii) in determining the intrinsic value of a life (or part of a life) for the one who lives it is to simply sum up all the values provided by CS(ii).

Before looking more closely at what this theory is, a quick note is in order about what this theory is *not*. Caring Satisfactionism is not a unified theory. In other words, CS will tell you how well the life of a person is going for the person living it, but it will not have anything to say about the lives of pigs, puppies, bats, gods, or fetuses. Having argued extensively against the possibility of a unified theory of welfare due to a variety of metaphysical and epistemological concerns, there is little reason to comment on this further here. Having said that, there is one type of non-person that is covered by CS. Recall that Frankfurt (1988: 16) distinguishes between persons and wantons. A

person cares about her will while a wanton does not. Frankfurt goes on to make a distinction within the class of wantons that is important for our purposes. A wanton either cannot or does not care about his first-order desires or which of them will or would move him to act. However, the difference between *cannot* and *does not* is substantial. If he cannot care about his will, then the wanton would be excluded from CS because of "his lack of the capacity for reflection" (Frankfurt 1988: 18-19). Lacking even the *capacity* for higher-order desires that serves as the basis for caring, this class of wantons is excluded from CS's ambit. The other class of wantons, however, is included within CS. This is the class whose lack of caring is due to their "mindless indifference to the enterprise of evaluating [their] own desires and motives" (Frankfurt 1988: 19). In other words, these are beings that *could* care about something, yet they do not. For such beings, CS will yield a personal welfare score of zero, which is the correct score because, as I have previously argued, there is no way to make the life of such a being go better or worse. Truly caring about nothing, there is no way to benefit or harm them.

II. Features Of Caring Satisfactionism

While I aim to demonstrate several of the advantages of CS in the remainder of this chapter, there is one in particular that I should like to make explicit at the outset: This is the *correct* theory of personal welfare. Making this statement here allows me the opportunity to claim, even though I will not have the space to demonstrate, that CS will provide the right answer in each scenario detailed previously where other theories have failed. It is also worth

mentioning this here because, obviously, all the other features of CS detailed below, while perhaps compelling and advantageous, would not be enough to make up for the fact that the theory is simply wrong.

Other than being correct, the two main advantages of CS are that it avoids both paternalism and bad metaphysics. Although it need not be the case, these two features can typically be found together. Objective-list theory, for instance, claims that certain things—whether the person cares about them or not—just do make a life go better (paternalism) and conjures up some fanciful objects in order to provide the necessary grounding (bad metaphysics). CS, on the other hand, avoids both of these pitfalls. First, CS allows for the fact that a good life for the person who lives it can be made up of anything at all and that this anything at all is determined exclusively by the person.[4] In

[4] Another way of demonstrating how CS avoids paternalism concerns is to examine how CS measures up to two of the welfare principles set out in Chapter One—the Internalist Principle' (IP': The value of a life (or part of a life) for the one who lives it is determined to a significant degree by what the person in question cares about.) and the Principle Concerning Paternalism (PCP: Paternalistic claims in axiology must be justified by a compelling theoretical interest and must be narrowly tailored to serve that interest.). CS clearly meets IP' by claiming that changes in welfare always involve what the person in question cares about. Similarly, CS meets PCP. The only aspect of CS that could possibly be construed as paternalistic is the belief requirement. However, the belief requirement does serve a compelling theoretical interest because, as noted earlier, it is very hard to see how a state of affairs that one is completely unaware of could directly impact one's welfare. Moreover, the belief requirement is narrowly tailored to serve this theoretical interest by requiring only belief and not more robust concepts like awareness or knowledge. These features come into even sharper focus when we examine CS in relation to the underlying reason for the concerns about paternalistic claims in the area of welfare—that we may end up with a theory that claimed I had a very good life when I hated every moment of it. CS, belief requirement and all, could never produce this outcome.

certain instances this might seem to be a regrettable feature of reality (e.g., it might be refreshing to find a moral component to personal welfare), but it is usually a comforting and benign feature of reality and is, in any case, ensured by the very nature of reality itself. This is because any paternalistic claim must be buttressed by highly speculative, and thus highly questionable, metaphysics. Just consider what sort of object would be required to make it *true* that just *these* things, and not *those* things, (where both classes are unconnected to what the person in question actually cares about) can increase personal welfare. It would be quite odd for such things to be part of the fabric of reality and for us to discover that they so existed. So, second, CS does not require a belief in any strange objects at all. CS only requires one to believe in beliefs, desires, and desires about desires. Bad metaphysics are avoided, as not believing in beliefs is self-defeating and not believing in desires is—while coherent—barely so.

CS also handles the impact of free will on personal welfare in a natural and seamless manner. While it is true that just about any theory could be adjusted to take free will into account, almost any attempt to do so will appear to be ad hoc, will assign an arbitrary value to free will, or both. An example from Ishtiyaque Haji's *Freedom and Value* will help illustrate this point. Haji (2009: 5) has a similar intuition about the value of freedom and thus argues "that on promising hedonistic views or preferentist views or whole-life satisfaction views of welfare, free positive 'life atoms'—atoms that contribute to welfare value—are better than otherwise similar unfree atoms, and free

negative life atoms are not as bad as otherwise similar unfree atoms." Haji (2009: 6) proceeds to advance considerations in support of the view that the correct theory of personal welfare "should adjust the value of its life atoms to reflect the extent to which they are free," defends "selected versions of these freedom-sensitive views against various objections," and then settles "on what I deem is the best contender." This best contender is illustrative for our purposes, as it claims that "free intrinsic attitudinal pleasures are better than otherwise similar unfree ones, and free intrinsic displeasures are not as bad as otherwise similar unfree ones" (Haji 2009: 6). Haji (2009: 26-28) arrives at this version of the theory after considering—and rejecting—a version that gives no weight at all to unfree pleasures and displeasures. He rejects this version after accepting the objection that, especially as it regards displeasures, unfree atoms are not intrinsically *worthless* and that freedom merely *enhances* the value of a life atom (Haji 2009: 27-28). *Given Haji's general approach*, it seems that he is clearly right to accept the objection that not all unfree atoms are intrinsically worthless, which then leads to his accepting that freedom merely enhances the value of a life atom.

However, the problem is with Haji's general approach, which is to consider the effects of *adding* free will to various established theories of personal welfare. Then, when he accepts the almost undeniable objection that unfree atoms are not *completely* worthless, especially as it pertains to pains or displeasures, he is forced to retreat to the position that freedom merely *enhances* the value of the life atoms of pleasure and displeasure. This stance

invites the inevitable question: *How much* does freedom enhance the value of these life atoms? Haji, understandably, does not offer an answer. This is because adding free will in such a seemingly ad hoc manner is going to force the value of free will in personal welfare calculations to seem entirely arbitrary.

CS, on the other hand, has free will built in, as expressed through what we care about, from the beginning. CS shows how free will impacts personal welfare, and no arbitrary value of free will is required in order to state the complete theory. It is hard to see how a modification of another theory could accomplish these things.

Finally, Caring Satisfactionism gives our reflective attitudes a central place in a theory of personal welfare. Indeed, it is precisely the failure of other theories to treat higher-order desires differently than first-order desires that has led to some counterintuitive results. This lack of differentiation may lead other theories to claim that a life filled with failures associated with what we care about most is still a good life due to a plethora of first-order desire satisfactions. Or that cases of weakness of will pertaining to very strong first-order desires may actually promote one's welfare more than self-control. In short, failing to properly account for reflective attitudes is a serious defect in other theories of welfare, whether it be desire theories or others.[5]

---

[5] The idea that the correct theory of welfare must give a central place to our reflective attitudes appears in many places in the literature. For example, Haybron (2008: 195) writes: "Pleasures or cravings that the individual does not, or would not, endorse on reflection cannot be allowed to trump the individual's own best judgment." In addition, Kraut (1994: 40), who ultimately rejects desire theory as we will soon see, writes that the most plausible form of a desire theory might give "special weight to second-order desires. The general

III. Caring Satisfactionism's Overlap With Other Theories

A theory of personal welfare is (usually) offered by someone because it has some degree of initial plausibility. For example, there is no theory that claims your life goes better for you to the extent you get purple things and worse for you to the extent you hear Christian rock. Accordingly, it tends to be a productive exercise to attempt to ascertain what actually gives each theory the degree of plausibility it has. Moreover, it is often thought to be a sign of increased plausibility of a theory if it can incorporate, and perhaps explain, these central intuitions of competing theories. As will hopefully become clear in the next few pages, CS nicely and naturally captures the central intuitions of many competing theories.

Sidgwick (1907: 111-12) discusses the claim that "a man's future good on the whole is what he would now desire and seek on the whole if all the consequences of all the different lines of conduct open to him were accurately foreseen and adequately realised in imagination at the present point of time." CS both incorporates this idea discussed by Sidgwick and expands upon it. Of course, CS incorporates the general idea that a person's future is what will ultimately complete his personal welfare picture (as it seems like any theory of welfare must), but it also supplies the *reason* to choose one path over another. One could imagine a person vividly foreseeing all the consequences of all the lines of conduct available to her—just as Sidgwick describes—and then asking,

---

idea is that so long as one wants something wholeheartedly and with open eyes, then it is good for one's desire to be satisfied, regardless of the content of the desire."

"Okay, what now? On what basis am I supposed to choose one of these over all the others?" The response from CS would be, roughly, that she should choose the one that allowed her to care deeply about a wide array of her strongest first-order desires, have those desires satisfied, and strongly believe those desires had been satisfied.

This idea is very much in line with what Rawls (1999: 370) says when he "trie[s] to fill in Sidgwick's notion of a person's good. In brief, our good is determined by the plan of life that we would adopt with full deliberative rationality if the future were accurately foreseen and adequately realized in imagination." While Rawls does admit that "from the definition above very little can be said about the content of a rational plan, or the particular activities that comprise it," (1999: 372), he does claim that in

> adjusting Sidgwick's notions to the choice of plans, we can say that the rational plan for a person is the one . . . that would be decided upon as the outcome of careful reflection in which the agent reviewed, in the light of all the relevant facts, what it would be like to carry out these plans and thereby ascertained the course of action that would best realize *his more fundamental desires*." (1999: 366) (emphasis added)

Rawls goes on to emphasize this point about the "good" for a person being the satisfaction of fundamental desires:

- "Someone is happy when his plans are going well, his more important aspirations being fulfilled, and he feels sure that his good fortune will endure" (Rawls 1999: 359).

- "The main features of a plan encourage and secure the fulfillment of the more permanent and general aims" (Rawls 1999: 360).

- "But I shall suppose that while rational principles can focus our judgments and set up guidelines for reflection, we must finally choose for ourselves in the sense that the choice often rests on our direct self-

knowledge not only of what things we want but also of how much we want them. Sometimes there is no way to avoid having to assess the relative intensity of our desires. Rational principles can help us to do this, but they cannot always determine these estimates in a routine fashion. To be sure, there is one formal principle that seems to provide a general answer. This is the principle to adopt that plan which maximizes the expected net balance of satisfaction. Or to express the criterion less hedonistically, if more loosely, one is directed to take that course most likely to realize one's most important aims. But this principle also fails to provide us with an explicit procedure for making up our minds. It is clearly left to the agent himself to decide what it is that he most wants and to judge the comparative importance of his several ends" (Rawls 1999: 365-66).

- "In this account of deliberative rationality I have assumed a certain competence on the part of the person deciding: he knows the general features of his wants and ends both present and future, and he is able to estimate the relative intensity of his desires, and to decide if necessary what he really wants" (Rawls 1999: 367).

From these quotes, it is not clear that Rawls would disagree with CS at all. At a minimum, CS incorporates Rawls's most basic intuitions about what constitutes a person's welfare.

CS also nicely encompasses the various forms of Aim Achievementism that have recently been propounded. The form of the theory that Simon Keller (2004: 36) finds defensible is this: "One aspect of an individual's welfare is her achieving her goals through her own efforts, regardless of what those goals are. It is not the only aspect of individual welfare, but it cannot be reduced to any of the others." T.M. Scanlon (1998: 124-25) makes a similar claim and then expounds on the fact that neither he nor Keller offers a complete theory:

> Leaving this question open, I conclude that any plausible theory of well-being would have to recognize at least the following fixed points. First, certain experiential states (such as various forms of satisfaction and enjoyment) contribute to well-being, but well-being is not determined solely by the quality of experience. Second, well-being depends to a large extent on a person's degree

> of success in achieving his or her main ends in life, provided that these are worth pursuing. This component of well-being reflects the fact that the life of a rational creature is something that is to be *lived* in an active sense—that is to say, shaped by his or her choices and reactions—and that well-being is therefore in large part a matter of how well this is done—of how well the ends are selected and how successfully they are pursued. Third, many goods that contribute to a person's well-being depend on the person's aims but go beyond the good of success in achieving those aims. These include such things as friendship, other valuable personal relations, and the achievement of various forms of excellence, such as in art or science.
>
> These intuitive fixed points provide the basis for rough judgments of comparative well-being: a person's well-being is certainly increased if her life is improved in one of the respects just mentioned while the others are held constant. But this list of fixed points does not amount to a *theory* of well-being. Such a theory would go beyond this list by doing such things as the following. It might provide a more unified account of what well-being is, on the basis of which one could see why the diverse things I have listed as contributing to well-being in fact do so. It might also provide a clearer account of the boundary of the concept—the line between contributions to one's well-being and things one has reason to pursue for other reasons. Finally, such a theory might provide a standard for making more exact comparisons of well-being—for deciding when, on balance, a person's well-being has been increased or decreased and by how much.

CS does provide a complete theory, and it encompasses the credible bits of each of these views. CS, in accordance with Keller's ideas, claims that willing yourself to accomplish goals that you care about will increase your welfare, and it puts this on equal footing with every other aspect of individual welfare (i.e., all the things a person cares about).[6] CS also incorporates much

---

[6] In this way, CS avoids the paternalism of Keller's view, which claims that there is added value, other things being equal, in bringing about a state of affairs yourself as opposed to its coming to obtain in another way. CS avoids this by recognizing that there is no compelling theoretical interest—as required by PCP—in claiming that it is better *for me* if I bring about X when I may not care how X obtains as long as it does.

of Scanlon's less permissive version of the theory. The "fixed points" he lists are all often what constitutes a person's "well-being", but there is a common problem for each of the three things he lists: What happens when the person does not care about one (or all) of these things? One can easily imagine people who do not care about anything beyond "experiential states," who do not care if their main ends are supposedly "worth pursuing," and who do not care about personal relations, art, science, etc.[7] CS provides a full theory where Scanlon does not by doing exactly what he says a full theory needs to do—going "beyond this list." CS provides a unified account of what well-being is for persons, provides a clear account of the boundary of the concept, and provides a standard for making more exact comparisons of well-being. Moreover, in accomplishing the things Scanlon requires of a theory, CS highlights just exactly where Scanlon's own disjointed, partial theory goes wrong.

Caring Satisfactionism also proves useful in understanding why most theorists prefer *global* desire-satisfaction theories to *summative* desire-satisfaction theories. These terms, apparently coined by Parfit (1984: 496-99), are misleading, however, since both types of theories are summative in nature. As used in this debate, summative theories appeal to all of a person's desires in determining welfare, whereas global theories omit "local" desires and appeal only to desires "about some part of one's life considered as a whole, or about

[7] As I have argued extensively for these conclusions previously, I will not rehash them here.

one's whole life" (Parfit 1984: 497). Parfit finds the global versions of the theories more plausible and gives two examples to help illustrate his point.

In his first example—let us call it Drug Addict—Parfit (1984: 497) knows that you subscribe to a summative theory and says he is going to make your life go better by making you a drug addict. Each morning you will awake with a very strong desire for a fix. Not to worry though—Parfit is going to keep you stocked up for the rest of your life, and the injection, the after-effects of the drug, and the desire for the drug will be neither pleasant nor painful. Everything else about your life stays the same other than your desire not to be addicted, which is a less intense desire than your daily desire for a morning fix. A summative theory will say that Parfit made your life go better, as being addicted to this drug increased your net desire satisfaction. Parfit rejects this conclusion, as the desires generated by the addiction—and the subsequent satisfaction thereof—are neither pleasant nor painful.

While the drug-addict case is a bit hard to follow and may not generate the same intuitions in everyone, Parfit's (1984: 498) next example—let's call it Great Life—is very compelling. Parfit asks us to imagine two possible lives. In the first life, Parfit (1984: 498) lives for fifty years, and these fifty years are of an extremely high quality: Parfit "would be very happy, would achieve great things, do much good, and love and be loved by many people." In the second life, Parfit (1984: 498) would live indefinitely, but the quality of his life would be such that it was always just barely worth living: "there would be nothing bad about this life, and it would each day contain a few small pleasures."

Summative theories will have to claim that the second life is better for Parfit if it lasts long enough. Parfit (1984: 498-99) disagrees; he thinks that, in both Drug Addict and Great Life, the first lives are better and, as a result, he rejects summative theories.[8]

Leaving the choice of lives in Drug Addict aside for the moment, it seems likely that almost everyone will choose the first (great) life in Parfit's Great Life example. If that is so, then it supports the conclusion that summative theories are less plausible than global theories.[9] A solid understanding of the reason for this should prove to be instructive. First, note that I claimed almost everyone would choose the fifty-year life in Great Life. This is an important qualification, and how we characterize the dissenters has significant implications. If we characterize the decision of the people who would choose what Parfit (1986: 160) calls elsewhere the "Drab Eternity" as being *necessarily* wrong, then Parfit's global theories are merely an attempt to put lipstick on the objective-list-theory pig. The claim that the dissenters in this case are necessarily wrong is equivalent to the claim that the things contained in the fifty-year life *just do* make a life go better no matter what anyone's attitude

---

[8] Carson (2000: 73-74) examines the same issue and agrees with Parfit that "the global view is preferable to the summative view." Griffin (1986: 144-45) also endorses the primacy of global desires.

[9] It is not my goal to present a full-fledged defense of global theories. For the purposes of this project, I think it is sufficient to show that summative theories get the wrong answer in, at a minimum, Great Life, and that global theories give us a way, at least on their face, to choose the fifty-year life in Great Life. In other words, global theories address desires about one's whole life or some part thereof and these theories, at least presumably, allow one to desire the fifty-year life over the indefinite.

toward them might be. This would be a claim that appeals to facts that are directly about value—a hallmark of objective-list theories. Desire-satisfaction theories like CS, on the other hand, are purely descriptive in that they claim that things only have value insofar as they involve desires. In other words, a thing makes a person's life go better *if and because* it is desired. As I reject claims that appeal directly to facts about value for the reasons previously stated and because I think there is a better solution to this issue, I will not treat Parfit's intuitions in Great Life as an endorsement of objective-list theories.

Parfit (1984: 498) claims that "there are countless cases in which it is true both (1) that, if someone's life went in one of two ways, this would produce a greater sum total of local desire-fulfillment, but (2) that the other alternative is what he would globally prefer, *whichever* way his actual life went." Drug Addict and Great Life are meant to be two of these countless cases, but if we do not appeal directly to facts about value, then what is the best way to explain them within a desire-satisfaction theory framework? Caring provides the explanation for most people's intuition in both cases and, I think, in any of the other countless cases as well. In the case in which Parfit (1984: 497) offers to make you a Drug Addict, he says "we can plausibly suppose that you would not welcome my proposal." The best explanation for this is that the average person does not care about becoming a drug addict and, perhaps, does care about *not* becoming a drug addict due to some second-hand experience with this topic.

Resorting to caring as a solution becomes even more obvious when we examine Great Life. In Parfit's (1984: 498) example, there is a great deal to care about in the fifty-year life for most people; you "would be very happy, would achieve great things, do much good, and love and be loved by many people." In the alternative life of infinite duration, nothing at all that most people care about occurs; "there would be nothing bad about this life, and it would each day contain a few small pleasures" (Parfit 1984: 498). Therefore, it is plausible to suppose that your personal welfare in Great Life could be very high, while in Drab Eternity it would be at or near zero—on par with, in terms of welfare, never having lived at all.[10]

Caring Satisfactionism can, I think, be employed to explain most of the central intuitions of most (all?) of the other theories of personal welfare. This goes, of course, for both of the other major theoretical camps on offer—hedonism (most people care about the attainment of pleasure and the avoidance of pain) and objective-list theories (many, or perhaps most, people care about getting many, or perhaps most, of the items that appear on most of the proposed objective lists). However, the time has come to determine if CS

---

[10] Considering a slight variant of these cases might also prove to be instructive. Leaving the details of Great Life as they are, we can change the details of Drab Eternity to directly challenge CS. In *Fab Eternity* there would again be nothing bad about this life, and it would each day contain the satisfaction of a few first-order desires that you care—only slightly—about having satisfied. Here CS will claim, rightly, that life goes better for you in Fab Eternity than it does in Great Life. The claim that Great Life is better for you than Fab Eternity (remember: there is *nothing bad* about this life and each day will contain things that *you actually care about*) is not plausible.

can handle two of the problems that are thought to be, if not deadly, then extremely difficult for any desire-based approach to solve.

IV. The Problem Of Self-Sacrifice

It is hard to overstate both how detrimental to a desire theory approach the problem of self-sacrifice is thought to be and how widely this belief is held among philosophers who work in this area. In what follows I will provide the necessary background for both of these claims and then demonstrate how CS solves the problem of self-sacrifice.

Before addressing the actual argument from self-sacrifice against desire theory, we should be very clear about both the target of the argument and the indispensable premise that drives it. The target of the argument is desire satisfactionism, of course, but a very particular form in which the agent's self-interest is whatever the agent wants most to do. The bedrock premise that drives the argument is that there are at least some cases of actual self-sacrifice in the world.

The argument, then, is as follows: (1) unless a theory of personal welfare can account for this premise (i.e., allow for the possibility of self-sacrifice), then the theory is false; (2) desire theory cannot allow for the possibility of self-sacrifice; (3) therefore, desire theory is false. The reason for desire theory's failure, or so the argument goes, involves the overlap between the target version of the theory and the components of self-sacrifice. In what appears to be the article in which this objection was first given a thorough treatment, Mark Carl Overvold (1980: 109-14) claims, roughly, that for an act to be one of

self-sacrifice, (1) the loss must be anticipated, (2) the act must be voluntary, and (3) there must be at least one other alternative open to the agent that would be more in his self-interest. For the purposes of clearly formulating the objection, it will help to reformulate (1), as Overvold himself does later (1980: 119), so the agent has "an accurate assessment of his alternatives" or "knows what he is doing." The problem is that the target version of the theory and the analysis of acts of self-sacrifice share (1) and (2), yet they reach different conclusions on occasion in (3). In other words, the target version of desire satisfactionism claims that if a person knows what he is doing and he does it voluntarily (where we only voluntarily do what we most want to do), then (3') the act just is what is in our self-interest. The analysis of acts of self-sacrifice claims that some acts that the agent knows he is doing and does voluntarily are (3) *not* what is in the agent's self-interest.

Although this objection was raised relatively recently, it was the inevitable next step in the debate over desire theory. When it was first suggested that well-being was increased when a person got what she wanted, the critics' obvious first move was to point to cases in which it was pretty clear that well-being was not enhanced when a desire was satisfied. As pointed out in an earlier chapter, the proponents' next move was to claim that if the agent has been properly informed about the consequences of her desire and then performed the act aimed at satisfying her desire, then the satisfaction of this *informed* desire would enhance her well-being. The argument from self-sacrifice is the critics' inevitable response to this form of the theory—the agent

is informed about the consequences of her act, she voluntarily acts to satisfy her desire, and it still does not enhance her well-being. And it is precisely due to the rise in popularity of *informed* desire versions of the theory that this has become such a popular objection. In addition to Overvold, this objection (or its nearest cousin—the accusation that desire theory has an incurable case of egoism) has been pressed by Sen (1977: 323), Brandt (1982: 173), Schwartz (1982: 199), Griffin (1986: 316 n.25), Haslett (1990: 79-80), Sumner (1996: 134), Adams (1999: 89), Carson (2000: 76), and Darwall (2002: 27).

Although this list of formidable philosophers makes this objection seem rather daunting, Caring Satisfactionism actually handles genuine cases of self-sacrifice quite well. Before demonstrating how CS handles these cases, however, we should be clear about what, exactly, these cases are. First, and most obviously, the act at issue must be *voluntary* (Overvold 1980: 109). It is not a case of *self*-sacrifice when the Fat Man is pushed onto the tracks to stop the trolley and its five passengers from suffering a fiery crash. Second, the agent must anticipate the loss (Overvold 1980: 109). If I voluntarily step in front of you and thereby take a bullet that I did not know was coming, then this is also not a case of self-sacrifice. Finally, we must distinguish between self-sacrifice and what Overvold (1980: 109) calls "cutting one's losses," which involves cases in which the agent "is forced to choose among his desires for incompatible things," yet the chosen act is not actually worse for the agent (Overvold 1980: 109). For example, one might elect to have surgery, and thus sacrifice some degree of short-term comfort, for the sake of one's long-term

health. While this does involve a sacrifice, it is not a case of self-sacrifice because the surgery is in the agent's best interests considered as a whole. Acts of self-sacrifice must actually involve embarking upon a path that is not in the agent's best interests considered as a whole.

To see how CS deals with these cases, an example from Overvold (1980: 115) will prove illuminating:

> Consider the following which we would normally regard as a central case of self-sacrifice: During a war, a man is ordered to carry out an operation which he knows will cause the death of many innocent victims. He correctly believes that he has only two viable alternatives: carry out the order and be rewarded for his efforts with fame and promotion, or refuse, and as a result lose his own life. Given a week to decide, he carefully considers these options, and finally refuses and as a result is executed. Since he knew he would lose his life, but still chose to perform the act, he meets the first two conditions of self-sacrifice. But did he actually suffer a loss of welfare? Ordinarily, we would treat his loss of life as sufficient for saying that there was a net loss of welfare, and hence that he had acted contrary to his self-interest. We can strengthen the claim if we add that the man correctly believed that although he would be bothered by guilt for awhile if he carried out the orders he would eventually get over it and live a happy and rewarding life.

The interesting thing about Overvold's description of the soldier's case is that he seems to anticipate the correct general form of desire theory that will avoid this objection without realizing it. In the last sentence of the excerpt, Overvold (1980: 115) says that "we can strengthen the claim" basically by stipulating that, if the man carries out the order, his welfare will be high in the *future*. Overvold, unfortunately, stops his inquiry here. The path for the desire theorist, however, should be clear. One need only ask: What would make it the case that the soldier's post-war welfare would be high? The answer from desire

theory generally—and CS in particular—is that there are a great number of relevant desire (or caring) satisfactions in the future part of his life.

The key to the Caring Satisfactionism solution to the problem is summing up and comparing all of the caring satisfactions and frustrations in both life paths available to the soldier. When the soldier is executed, there are no future caring satisfactions or frustrations—the story of how well his life went for him is over. When the soldier lives, his life will include all of the caring satisfactions and frustrations contained in the life in which he is executed (at least up to the point he makes his decision) *plus* the net positive balance of caring satisfactions and frustrations he enjoys in his post-war life (i.e., his "happy and rewarding life"). In this way, CS allows for cases of self-sacrifice and correctly tells us that this is a genuine case.

Well that's all well and good, or so you might say, but what was all the fuss about then? The sheer number of philosophers throwing self-sacrifice rocks at desire theory's windows does seem to cry out for an error theory. There is much that could be written here, but I shall attempt to be brief. First, the version of desire theory at which Overvold's influential formulation of the objection is aimed was a version formulated by Brandt (1972: 686-86). Brandt's version is an informed desire account (which I argued against thoroughly in Chapter Four) and which is so imperfectly formulated that Brandt (1979: 247) has himself rejected it. Second, proponents of this objection from self-sacrifice, for the most part, have their own theories of personal welfare and thus may not be trying to reformulate desire theory in

order to meet this objection. Finally, there appears to be a very prevalent misconception about desire theory that is doing most of the work here. This misconception is that desire theory must (often? usually?) claim that if I prefer one path to another, then the path I prefer, and thus choose, is the path that maximizes my welfare. This is not a claim that desire theory needs to make, and it should not do so since the claim is quite clearly erroneous. If self-sacrifice is possible—and it is—then we do not always choose what is best for us, even assuming, as informed desire theories do, that we know everything that will occur as a result of any action under consideration.[11] Even though desire theory might claim that the satisfaction of the initial desire to embark on the sub-optimal path is good for the agent, it need not claim that the *rest* of the path is best for the agent. This is because the theory need not rely on the agent's choice to determine the best path. Rather, as suggested earlier, the

---

[11] For an interesting account of knowing what is best and not choosing it, see Dostoevsky's (1989: 3) Underground Man in *Notes from Underground*:

> I am a sick man. . . . I am a spiteful man. I am a most unpleasant man. I think my liver is diseased. Then again, I don't know a thing about my illness; I'm not even sure what hurts. I'm not being treated and never have been, though I respect both medicine and doctors. Besides, I'm extremely superstitious—well at least enough to respect medicine. (I'm sufficiently educated not to be superstitious; but I am, anyway.) No, gentlemen, it's out of spite that I don't wish to be treated. Now then, that's something you probably won't understand. Well, I do. Of course, I won't really be able to explain to you precisely who will be hurt by my spite in this case; I know perfectly well that I can't possibly "get even" with doctors by refusing their treatment; I know better than anyone that all this is going to hurt me alone, and no one else. Even so, if I refuse to be treated, it's out of spite. My liver hurts? Good, let it hurt even more!

theory can simply sum the relevant caring satisfactions and frustrations on each path in order to determine what is best.

V. Desiring The Bad & Desiring Not To Be Well-Off

The desire for the bad and the desire not to be well-off are sometimes grouped together and presented as one objection, but they actually pose different problems for desire theories. In this section I will examine these desires and their related problems to see if they prove problematic for Caring Satisfactionism.

One of the problems associated with the desire not to be well-off is in the form of a paradox. For example, suppose I have just two desires: I want a lover who won't drive me crazy with an intensity of 5, and I want not to be well-off with an intensity of 10. Employing a very basic desire-satisfaction theory of welfare, when my lover drives me crazy (thus frustrating my desire), my welfare score is a -5. So now I am not well-off. However, my desire not to be well-off is now satisfied, which seems to give me a welfare score of +5. But now I am no longer not well-off. As Heathwood (2005: 502) says, "My desire that my welfare be negative is satisfied if and only if it is not satisfied." This paradox occurs only when welfare hovers around zero, but it is a paradox nonetheless.

Feldman (2004: 17) and Bradley (2007: 46) both think this paradox makes desire theory at least less plausible and possibly false. However, Heathwood (2005: 502-03) points out that even if the desire not to be well-off is possible, this paradox is simply inherited from paradoxes involving desires generally and that "however the more basic paradoxes are solved so will the

paradoxes for desire-satisfaction theories be solved." In other words, this paradox is everyone's paradox, not just the desire theorist's.[12] Since this is quite clearly the case, this problem need not detain us any longer.

The other problem often associated with the desire not to be well-off is, perhaps, more worthy of an in-depth analysis. Desire theory claims, roughly, that things I want are good for me if and because I want them. This version of the desire-not-to-be-well-off objection says that sometimes a person can want things *because* they are bad for him (Adams 1999: 89). If that is true, then these bad things are not good for me by definition and, therefore, desire theory is false.

The first task presented by this objection is to attempt to understand what it means. The most plausible versions of desire theory, including CS, claim that, roughly, having a desire satisfied is always good for you even though it may be *all-things-considered* bad for you because of what the satisfaction of your desire leads to. For example, the satisfaction of my desire for a drink is good for me even though it causes a state of affairs that is all-

---

[12] Moreover, Bradley (2007: 47) notes that there is a paradox of the type described lurking for every theory that claims "how well things go for a person is determined by (i) the person's attitudes towards states of affairs or propositions (desiring them, believing them, taking pleasure in them), and (ii) whether those states of affairs are true." In other words, the only theories that can avoid this sort of paradox either completely ignore the person's attitudes toward the things that are supposedly good for them (e.g., objective-list theory), or completely ignore whether the state of affairs the person has the relevant attitude toward actually obtains (e.g., hedonism of the experience-machine-lives-are-great-for-welfare variety), or completely ignore both of these things (e.g., I have no idea what such a theory would look like). Accordingly—if the analysis was correct up to the beginning of Chapter Five—then a paradox of this sort *is* going to arise for the right theory of welfare.

things-considered bad for me because the water was polluted and makes me sick. This objection is aimed at eliminating the idea that the initial desire satisfaction is always good for a person. And it does this by appealing to another theory of welfare without expressly doing so and without naming it. Remember the case: You want something *because* it is bad for you. The only theory that can do the intuitive and theoretical work here is some version of an objective-list theory.[13] In other words, the objection claims that there just are some sensations, or mental states, or *something* that is bad for a person to get, that it is possible to want these things for their bad qualities, and that therefore desire theory is false.

This version of the desire-not-to-be-well-off objection results in a solid strategy for the objective-list theorist because it is at precisely this point that objective-list theory in general, and hedonism in particular, is most plausible. For it does, at first blush, appear that certain things like pain, or boredom, or NASCAR just are not good for you whether you want them or not. However, as we saw previously, objective-list theories are not plausible and, accordingly, should not be able to get the thin edge of their wedge in using this objection.

Although this objection is little more than an attempt to reopen the objective-list theory debate on terms favorable to that family of theories, let us suppose that I am wrong here. What should desire theory generally, and CS in

---

[13] For purposes of simplicity here, I am grouping hedonism in with other objective-list theories since hedonism is just a (very short) objective-list theory. Of course, there could be a version of "hedonism" that claimed that pleasure and pain were a function of our desires, but this version of hedonism could not serve as the basis for this objection.

particular, have to say about the desire not to be well-off? The first thing to note here is the difficulty in assessing this scenario due to the murkiness of the human psyche. Nowhere is this fact more obvious than it is here, as what we are being asked to consider and evaluate is a condition in which, roughly, a person wants something because he does not want it.[14]

Described in this way, this objection should immediately raise doubts as to the possibility of this desire. Perhaps this doubt is the reason this objection has not been pursued more frequently in the literature. In fact, there is only one supposed case in any of the literature (to which everyone else cites), and it is not clear that this one case demonstrates what it purports to. The belle of the desire-not-to-be-well-off ball is Richard Kraut's (1994: 40-41) self-punisher:

> It is conceptually and psychologically possible for people to decide, voluntarily and with due deliberation, to renounce their good in favor of an alternative goal. They can clearheadedly design a long-range plan and fulfill it, thereby satisfying their deepest desires, in spite of the fact that they realize all the while that what they are doing is bad for them. In fact, they can carry out certain plans precisely because they think that it is bad for them to do so. For example, suppose a man has committed a serious crime at an earlier point in his life, and although he now regrets having done so, he realizes that no one will believe him if he confesses. So he decides to inflict a punishment upon himself for a period of several years. He abandons his current line of work, which he loves, and takes a job that he considers boring, arduous, and insignificant. He does not regard this as a way of serving others, because he realizes that what he will be doing is useless. His aim is simply to balance the evil he has done to others with a comparable evil for himself. Taking a pill to relieve his pangs of guilt would be of no use, since his aim is to do himself harm, not to make himself feel

[14] Assuming both that the Motivational Theory of Pleasure (which I argued for previously) is true and that this objection is not merely an objective-list theory in disguise.

> good. He punishes himself because he regards this as a moral necessity, and when he carries out his punishment, he does so from a sense of duty rather than a joyful love of justice and certainly with no relish for the particular job he is doing.

Feldman (2004: 17) and Adams (1999: 89) cite Kraut approvingly, but do not elaborate or provide additional detail. That is the entire list of those philosophers who seem to agree unequivocally with this particular objection and the subsequent conclusion.

Carson (2000: 88-92) and Heathwood (2005: 501-02) are the other two philosophers who discuss this version of the objection. Although they each view the case differently, neither agrees with Kraut's conclusion on the basis of the described case, and together they cover the most basic desire theory responses to this version of the desire-not-to be-well-off objection.

Heathwood (2005: 501-02) appears to take Kraut's self-punisher—let us call him Special K—at face value. Heathwood assumes that Special K has a desire not to be well-off and that taking the boring, arduous, insignificant job will most likely serve to make Special K's life go badly for him. This is of course contrary to Kraut's conclusion about what a desire theory would have to conclude about Special K. Heathwood (2005: 502) reaches his conclusion that Special K is probably not well-off by noting that his job is, by Kraut's own description, full of frequent, daily desire frustrations (e.g., to be bored is to want to do something else) and that all these frustrations add up to a bad life for the one who lives it. Granted, Special K does have one thing that is intrinsically good for him—the satisfaction of the desire not to be well-off—yet

this desire satisfaction probably does (and actually must)[15] count for less in terms of welfare than the sum of all the desire frustrations from his bad job. Thus, Heathwood uses desire theory to arrive at the conclusion Kraut claims desire theory could not produce—namely, that Special K's life is going badly for him.

Carson (2000: 89), on the other hand, says that "on the basis of Kraut's description of the case, I'm not sure that we should say that self-punishment is contrary to the person's long-term interests." Indeed, most instances of punishment, and perhaps all instances of morally justifiable punishment, are not aimed at making the punishee worse off long-term. Parents punish children millions of times per day, and the goal is rarely, if ever, to make the child's life go worse long-term. And if we relocate the punisher and the punishee within one psyche, it becomes less clear that the punisher could intend to make the punishee worse off long-term. As Carson (2000: 90) says, "In order to make this argument work, Kraut needs to describe the kind of case he has in mind very carefully."

Accordingly, even though Kraut (1994: 40) claims that it is "conceptually and psychologically possible" to carry out a plan because it is bad for you, it is not clear that this is indeed true. Moreover, Kraut's Special K case does not establish this as a legitimate possibility. Instead of describing Special K as

---

[15] The desire not to be well-off must count for less, in terms of welfare, than the total of all the frustrations; otherwise, the desire not to be well-off will no longer be satisfied. This is the paradox involving the desire not to be well-off that is discussed at the beginning of this section.

having a desire not to be well-off, we could describe him as having a desire to have a life that includes certain frustrations. However, this description is perfectly compatible with having a desire to be as well-off as he can given his past, and even with having a desire to be well-off, period. This is because it is reasonable to assume that Special K's feelings of guilt will ruin the first-order desire satisfaction he receives in the job he loves. He does not think he deserves them, has a higher-order desire to no longer have these first-order desires satisfied, and thus finds his life going worse as a result. Alternatively, if he takes the bad job, the feelings of guilt are easier to handle due to the frustration of many of his first-order desires for which he has a higher-order desire. This view of Special K's available options demonstrates (1) that the upper end of lives in terms of personal welfare is foreclosed to Special K due to his feelings of guilt, and (2) given that how well a life is going needs to be considered in relation to possible alternatives at the time, Special K's life goes best for him when he takes the bad job. He chose a course that is better all things considered, yet is worse from the standpoint of first-order desire satisfaction and frustration alone because he values the higher-order desire—one based on moral considerations with the object being certain first-order desire frustrations—in a way that he does not value the first-order desires. In this way, perhaps Special K's life is going well for him despite the bad job. And this might continue to be the case if and until his relevant mix of desires and their associated intensities change.[16]

---

[16] For much of this paragraph I am indebted to Michael Tooley's comments on

So Kraut's case, the only one in the literature, is not clearly a case of desiring the bad because it is bad, and it is not clear that this is even conceptually or psychologically possible. Moreover, depending on how one fills in the details, desire theory is quite capable of producing the answer that Special K's life is either going poorly for him or going as well as possible given his particular history and psychological make-up. The desire at issue is odd, at best, and it makes sorting out the relevant psychological features even harder than normal.

Caring Satisfactionism is also capable of producing a range of welfare scores depending on how we describe this type of case (assuming, again, that these desires are possible). Following Heathwood's reading of Kraut's Special K case, CS would accurately produce a poor welfare score for Special K. The basic idea is that Special K cares about not performing the "boring, arduous, and insignificant" tasks that fill his days. When he does perform these tasks, the desires Special K cares about are frustrated, and CS claims that his life is going poorly.

Alternatively, CS can produce the outcome that Special K's life is going as well for him as possible given the circumstances. Following Carson (2000: 90), let us assume that Special K "has an interest in his own moral purity." Accordingly, what Special K does care about is that some of his first-order desires should be frustrated. So when he takes the bad job and experiences these frustrations, he is as contented with this state of affairs as possible given

---

a prior presentation of these ideas.

his past history and subsequent regret because he takes these frustrations to be serving a moral function that he holds dear.

A few notes about this Carson-style reading of Kraut's case are in order. First, there is nothing odd about caring about having certain first-order desires frustrated. We often would like some of our first-order desires frustrated when they pertain to smoking, drinking, eating, etc. If there is fault to be found with the frustrations Special K seeks, then it will have to be because of Special K's underlying *reason* for caring about these frustrations. Since the reason is a moral one, we would have to exclude moral reasons from our theory of welfare. However, as I argued previously, this is a non-starter. Kraut (1994: 50 n.5) raises and rejects this very possibility: "Desires to be a good friend, a good father, or a good citizen all have moral content; and it is hard to see why we should rule out the possibility that satisfying these desires can be good." Second, it is probably correct to point out that Special K's life could be going much better for him if only he cared about other things. However, the problem here is Special K and his particular psychological make-up, not the theory. And if it were a problem with the theory, then it is a problem for every theory that gives any weight to the particular make-up of the person at issue (which, of course, any plausible theory does). Third, it is useful to remember that what Special K cares about could change at any time. So if he wakes up one day and thinks he has paid his dues and now cares about performing his old, good job instead, then CS will say that his life is now going poorly if he stays in the bad job and would go well if he were to get his old job back.

The bottom line for determining how Kraut's case should come out is getting to the bottom of what Special K really cares about. This is often difficult to determine, even for the person involved. The task is further complicated in this case because we are on the outside looking in at a fictional case with limited details, and we are supposed to try to understand a person's wanting what he does not want. However, as we have seen, CS is flexible enough to produce the correct outcome in this case, whichever description turns out to be accurate.

VI. The Paradox Of Welfare

Personal welfare, in terms of topics that have received serious, sustained philosophical treatment, is relatively young. Prior to 1980 or so, there is little more than a smattering of terse pronouncements from most of the leading figures in the history of philosophy, from Plato all the way up to Rawls. Given this somewhat surprising fact, it is little wonder, then, that some of the debate is still confused. Unfortunately, the debate is still deeply confused on at least two topics in particular. The first we have visited many times and in many ways over the course of this project. The issue involves the concept of personal welfare itself. The easiest, albeit rough, way to characterize this confusion is to describe it as a debate between those theorists who are attempting to define what constitutes a *good life* and those theorists who are attempting to define what constitutes a good life *for the person who lives it*. It is the latter category that is (hopefully) the topic of this project and what is meant by the terms and phrases well-being, personal welfare, a life that is intrinsically valuable for a

subject, what makes a life go best for the person who lives it, etc. However, it is the issue of what constitutes a good life, *full stop* (i.e., not for a subject but, perhaps, for the world) that many who are purportedly engaged in this debate are concerned with, whether or not they realize it. Objective-list theorists, perfectionism theorists, and those who argue for the plausibility of posthumous benefits and harms are just a few examples. For the most part, the theorists who think they are making a case for these things in the *personal welfare* debate either are confused about the topic under discussion or refuse to acknowledge that a good life for the person who lives it need not be pretty, interesting, or moral.

The second serious confusion is located entirely within the topic of personal welfare and, perhaps for this reason, is harder to spot. And I think it has done more to hamper progress than any other single issue in the debate. But before attempting to properly formulate the issue, I would like to put it in context in order to see where the difficulty arose.

Chapter Three of this project is a rather extensive treatment of, among other things, the origins of our desires and the role they play in the complexities of the human psyche. Perhaps this discussion seemed out of place in a project on well-being, yet its function was to combat a specific family of problems in the debate. This family of problems stems from the fact that within the *personal* welfare debate, rarely, if ever, is one presented with a nuanced and detailed description of the relevant parts of the human psyche in general and desires in particular. Reading the literature gives one the

impression that desires are the sort of thing you might find under your couch cushions or in the bulk section at Whole Foods. Of course it is not quite that bad, but desires are often presented as just these free-floating things inside your head.

This impoverished presentation of the mind and the desires within it has led to a series of problems for friends and foes alike in the debate over desire theory. Although Derek Parfit is undoubtedly a genius, and his *Reasons and Persons* (1984) is certainly brilliant and arguably the most influential single work in ethics of the twentieth century, the problem seems to have taken root there. Tucked away in Appendix I (yes, the ninth of ten appendices) of Parfit's tome is this (now extensively cited) story that we encountered previously:[17]

> Consider next Desire-Fulfilment Theories. The simplest is the *Unrestricted* Theory. This claims that what is best for someone is what would best fulfil *all* of his desires, throughout his life. Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfiment [sic] theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory. (Parfit 1984: 494)

Parfit's point is well taken. However, whereas I claim that the problem here is either that Parfit does not *really* care about the stranger or that Parfit never *knows* the stranger has been cured (or both), Parfit sends the debate in a different direction. "Another theory," Parfit (1984: 494) says in the next sentence after the quoted extract above, "appeals only to our desires about our

[17] Parfit was not the first to tackle this idea; that seems to have been Overvold in 1982. Yet the popularity of *Reasons and Persons* seems to have made Parfit's Stranger the common point of departure for discussions of this issue.

own lives." Parfit (1984: 494) clearly prefers this option, what he calls "Success Theory," to unrestricted desire-fulfillment theory, and he never explicitly rejects it. However, Parfit (1984: 494) does note that "when [success theory] appeals only to desires that are about our own lives, it may be unclear what this excludes." Parfit never formulates a clear answer to this. Instead, he considers (and rejects) the possibility that the desire that all of my desires be satisfied is about my own life, gives one example of a desire that is about my own life, one example of a desire that is not, and moves on.

Parfit's Stranger objection has moved the debate, I think, in one general direction that has manifested itself in at least three distinct objections. The first, and most general, way to formulate the objection is that desire theory is too broad. Griffin (1986: 16-17), citing Parfit's Stranger, phrases it just this way:

> The breadth of the [desire theory] account, which is its attraction, is also its great flaw. . . . The trouble is that one's desires spread themselves so widely over the world that their objects extend far outside the bound of what, with any plausibility, one could take as touching one's own well-being.

Kagan appears to be making a similar argument for the same idea with his Prime Number Fan. He asks us to imagine that he is a huge fan of prime numbers who desires that the total number of atoms in the universe is prime (Kagan 1998: 37). When this turns out to be the case, the version of desire theory he presents will claim that his life is going better as a result: "But this is

absurd. The number of atoms in the universe has nothing at all to do with the quality of my life" (Kagan 1998: 37).[18]

The second way of framing the objection is to offer it up as a necessary restriction designed to make desire theory more plausible by limiting the breadth of the account to desires that are about one's own life. As we have seen, this is what Parfit (1984: 494) proposes. Overvold, (1982: 188) writing before Parfit, suggested a similar restriction: "only desires and aversions that are relevant to the determination of an individual's, S's, self-interest are S's desires and aversions for states of affairs in which S is an essential constituent." Carson (2000: 76) has some reservations about Overvold's theory, but says that "nevertheless, to date, it is the best and most fully developed attempt to restrict the desire-satisfaction theory, and it is a theory that lends itself to further refinement and improvement." Portmore (2007: 27), citing both Parfit's Stranger and Kagan's Prime Number Fan, also thinks desire theory needs this restriction if it is to be plausible (which, apparently, he ultimately does not think it is) and claims the following:

> First, as most desire theorists acknowledge, the theory must be restricted in such a way that only those desires that pertain to one's own life count in determining one's welfare. The problem is that no one has yet provided a plausible account of which desires these are . . . .

---

[18] Again here, as with Parfit's Stranger, what is doing the intuitive work is the fact that Kagan could never *know* the number of atoms in the universe is prime, and it is hard to imagine anyone actually *caring* about this at all (although, as mentioned previously, it is obviously *possible* for a person to care about this).

Ultimately, this line of thought paints itself into a corner from which there is no escape. This can be seen in the third and final way this Parfit's-Stranger-style objection manifests itself. What began as an objection to the breadth of desire theory, and then focused on requiring desire theory to count only those desires that are about one's own life, now focuses even more tightly on needing to be *selfish*. As with many topics, Sidgwick (1907: 109) anticipates this move when he proposes to "consider only what a man desires for itself—not as a means to an ulterior result,—and for himself—not benevolently for others: his own Good and ultimate Good." Schwartz (1982: 199), also writing before Parfit's Stranger, takes the same line: "Roughly speaking, self-regarding preferences are ones not based on any ultimate objective of promoting the welfare, the goals, or the happiness of anyone but their subject. Only such preferences (and perhaps not even they) constitute strong evidence of what is good for this subject." Sen (1985: 190) pursues a similar line of thought when he claims "that connection [between what a person regards as valuable and the value of the person's well-being] will be weakened precisely by the fact that a person may well value things other than personal well-being." Sumner (1996: 120) seems to be getting at this same idea: "The fact that I choose x rather than y (or would choose it if I could) does not show that I expect a higher personal payoff from it, since my choice may be motivated by other considerations, such as altruism or a sense of obligation." Adams (1999: 87-88) continues this trend by claiming that both altruistic and idealistic

desires (involving virtues and other ideals) can lead you to do what is not best for you.

Thus we have the predictable end to this line of reasoning. Employing some misleading hypotheticals, desire theory is thought to be too broad. Given the nature of the misleading hypotheticals, the consensus becomes that desires relevant to welfare must be limited to those desires that are about your own life. But what does *that* mean? Can't it just be that *I* want whatever it is that I want? No, because then we are right back where we started—an account of welfare that is too broad because we can want anything. Well, what then? The desires that are relevant to welfare must be about your own life in that they must be selfish desires dealing directly with your own welfare.

The problem is that this road is a dead end. Accordingly, it is not a surprise to find that no one has been able to formulate a decent account of what welfare-related desires are at the end of this path. The reason for this is that the dead end actually takes the form of the paradox of welfare. Joel Feinberg's (2006: 531-32) description, although discussing the paradox of hedonism, illustrates the problem nicely:

> Imagine a person (let's call him "Jones") who is, first of all, devoid of intellectual curiosity. He has no desire to acquire any kind of knowledge for its own sake, and thus is utterly indifferent to questions of science, mathematics, and philosophy. Imagine further that the beauties of nature leave Jones cold: he is unimpressed by the autumn foliage, the snowcapped mountains, and the rolling oceans. Long walks in the country on spring mornings and skiing forays in the winter are to him equally a bore. Moreover, let us suppose that Jones can find no appeal in art. Novels are dull, poetry a pain, paintings nonsense and music just noise. Suppose further that Jones has neither the participant's

> nor the spectator's passion for baseball, football, tennis, or any other sport. Swimming to him is a cruel aquatic form of calisthenics, the sun only a cause of sunburn. Dancing is coeducational idiocy, conversation a waste of time, the other sex an unappealing mystery. Politics is a fraud, religion mere superstition; and the misery of millions of underprivileged human beings is nothing to be concerned with or excited about. Suppose finally that Jones has no talent for any kind of handicraft, industry, or commerce, and that he does not regret that fact.
>
> What then is Jones interested in? He must desire something. To be sure, he does. Jones has an overwhelming passion for, a complete preoccupation with, his own [welfare]. The one exclusive desire of his life is to [increase his own personal welfare]. It takes little imagination at this point to see that Jones's one desire is bound to be frustrated. People who—like Jones—most hotly pursue their own [welfare] are the least likely to find it. [People leading lives high in welfare] are those who successfully pursue such things as aesthetic or religious experience, self-expression, service to others, victory in competitions, knowledge, power, and so on. If none of these things in themselves and for their own sakes mean anything to a person, if they are valued at all then only as a means to one's own [personal welfare]—then that [welfare] can never come. The way to achieve [a good life for you] is to pursue something else.

The bracketed changes in the second paragraph of the excerpt were made to adapt Jones's case from illustrating the paradox of hedonism to illustrating the paradox of welfare. So instead of pleasure and happiness in the hedonism case, we now have welfare and good lives in the welfare case. However, once this change is made, the problem becomes obvious. If desire theory is limited to desires (which actually must reduce to the one desire) about your own welfare, then there is nothing left to produce the desired welfare. Thus the paradox of welfare: To increase your welfare, you must desire something besides your own welfare. So the "about you" aspect of desires need be nothing more than that *you* truly care about something as an end in itself.

The two best desire theorists of all time, Mark Carl Overvold and Chris Heathwood, seem to have come very close to pointing out this paradox. Although neither of them did, I think the tension between various positions they have taken would have eventually led them here. Notice the tension between the following two excerpts from Overvold. First, Overvold (1982: 186-87) states:

> Despite its popularity, there are difficulties with the prevailing account. The problems emerge when we reflect on the place of the concept of self-interest in a wider range of concepts including the concepts of self-sacrifice and selfishness. Consider an apparent case of self-sacrifice: An individual dies in an effort to save another. Now it seems to me that it would be not only incorrect, but unintelligible, to describe such a case as both a genuine instance of self-sacrifice and an act that enhances the individual's welfare. Let this be our first constraint on the concept of self-interest: The account of self-interest must not be so broad as to allow us to describe the same act as a self-sacrifice and as an act that promotes the agent's self-interest.

One paragraph later, Overvold (1982: 187) writes:

> But we must be careful not to restrict the concept of self-interest too narrowly. In this respect, the concept of selfishness is instructive. It provides our second constraint: For an account of self-interest to be adequate, the following proposition must be intelligible: Action A is unselfish, and action A maximizes my self-interest. This stipulation resists any attempt to restrict a person's self-interest to the class of actions that are performed for selfish motives. It has often been argued, most forcefully by Bishop Butler, that a life of caring for others, i.e., acting at least sometimes from unselfish motives, can in the long run maximize an individual's welfare. Although this need not be true, it does pose an intelligible possibility. An adequate analysis of the concept of self-interest must not make such an alternative conceptually impossible.

While most philosophers seem impressed with Overvold's attempt[19] to navigate the waters between self-sacrifice and unselfishness, I am not aware of any who accept it. Given the task Overvold set for himself, this is not surprising.

Heathwood has a similar tension in his writing on this topic. Heathwood (26) recites the following case from Gibbard (1987: 137):

> Consider a piece of cake to be divided between Desdemona and Iago. Desdemona is altruistic: given the choice, she would divide the cake equally. Iago is selfish, and given the choice he would take the entire cake for himself. They are of similar size and appetite, they eat cake with equal signs of gusto, and they will each undergo similar kinds of inconvenience in order to eat a cake that would otherwise go to waste.

Heathwood (26) then says:

> The problem for rationalistically-oriented desire theories of welfare is that they will imply that it would be best for Desdemona to get only half the cake, for this is the outcome she prefers considering the matter rationalistically. But intuitively what would be best *for her* would be to get all of it.

However, writing about the problem of self-sacrifice, Heathwood (2011a: 32) provides the following case:

> *Alice's Friday Night.* Alice is deliberating over how to spend her Friday night. She can go to the disco with her friends, or she can volunteer at the soup kitchen. Alice considers the options and, despite how badly she wants to go dancing with her friends, she decides, voluntarily and with full and vivid knowledge, to spend her Friday night helping the needy at the soup kitchen. She feels it would be the right thing to do, and so she does it.

Heathwood (2011a: 35) then has this to say about the case:

> [I]t is no longer very intuitive that Alice is doing what is worse for her on this Friday night. Indeed, the claim appears rather question-begging. Just because Alice isn't acting with *herself* in

[19] This attempt is described above as his attempt to formulate an analysis of desires that are about our own lives.

> mind, or with her own best interests in mind, and is instead acting benevolently for others, we cannot conclude that she must therefore fail to be doing what is in her best interests.

Just as Overvold had difficulty drawing a bright line between self-sacrifice and unselfish acts in the abstract, so too, I think, will Heathwood have difficulty distinguishing between Desdemona's self-sacrificial act and Alice's unselfish one.

Frankfurt, once again, brings the solution here into sharp focus. Although he is discussing "love" in this excerpt, the same could be said for the "caring" that I have been discussing throughout this project.

> The appearance of conflict between pursuing one's own interests and being selflessly devoted to the interests of another is dispelled once we appreciate that what serves the self-interest of the lover is nothing other than his selflessness. It is only if his love is genuine, needless to say, that it can have the importance for him that loving entails. Therefore, insofar as loving is important to him, maintaining the volitional attitudes that constitute loving must be important to him. Now those attitudes consist essentially in caring selflessly about the well-being of a beloved. There is no loving without that. Accordingly, the benefit of loving accrues to a person only to the extent that he cares about his beloved disinterestedly, and not for the sake of any benefit that he may derive either from the beloved or from loving it. He cannot hope to fulfill his own interest in loving unless he puts aside his personal needs and ambitions and dedicates himself to the interests of another.
>
> Any suspicion that this would require an implausibly high-minded readiness for self-sacrifice can be allayed by the recognition that, in the very nature of the case, a lover *identifies himself* with what he loves. In virtue of this identification, protecting the interests of his beloved is necessarily among the lover's own interests. The interests of his beloved are not actually *other* than his at all. They are his interests too. Far from being austerely detached from the fortunes of what he loves, he is personally affected by them. The fact that he cares about his beloved as he does means that his life is enhanced when its interests prevail and that he is harmed when those interests are defeated. The lover is *invested* in his beloved: he profits by its

> successes, and its failures cause him to suffer. To the extent that he invests himself in what he loves, and in that way identifies with it, its interests are identical with his own. It is hardly surprising, then, that for the lover selflessness and self-interest coincide. (Frankfurt 2004: 61-62)

While there are excellent attempts at solving the problem that began in earnest with Parfit's Stranger, there is no successful end game for the desire theorist on this path. The only way out—and the right answer—is the claim that all of these desires, and more, are the very stuff of welfare. As Frankfurt helps to illustrate, provided that I really care about serving the ends I have chosen, then satisfying these desires does make my life go better *for me*. After acknowledging that a path other than the one I have chosen may make my life go better for me, there is nothing else for the desire theorist to do.

Accordingly, there is not much left for me to do. Perhaps it is an odd feature of reality that in order to have a good life, we must care about something other than having a good life. Perhaps it is even odder still that in order to have a good life, we must open ourselves up to the possibility of having a bad life or, worse still, the possibility of having a life that is worse than never having lived at all. As Frankfurt (1999: 111) says, "A person who cares about something is, as it were, invested in it. By caring about it, he makes himself susceptible to benefits and vulnerable to losses depending upon whether what he cares about flourishes or is diminished." This is precisely why the attempts by some theories of welfare to claim the "flourishing" and disown the "diminishing"—as illustrated in the excerpts from Overvold and Heathwood—are doomed. Actual caring does not work this way, and caring is

the stuff of welfare. Frankfurt (2004: 23), entirely deserving of the last word, sums this up perfectly:

> It is by caring about things that we infuse the world with importance. This provides us with stable ambitions and concerns; it marks our interests and our goals. The importance that our caring creates for us defines the framework of standards and aims in terms of which we endeavor to conduct our lives. A person who cares about something is guided, as his attitudes and his actions are shaped, by his continuing interest in it. Insofar as he does care about certain things, this determines how he thinks it important for him to conduct his life. The totality of the various things that a person cares about—together with his ordering of how important to him they are—effectively specifies his answer to the question of how to live.

EPILOGUE: THE MEANING OF LIFE

Imagine a world exactly like the one in which we find ourselves, albeit with one difference. This new world comes complete with meaning labels. *Everything* has one. Parents, siblings, strangers, cars, buildings, plants, animals, places, events, ideas, art—they all have a meaning label. In addition, every person has a copy of the Holy Metameaning Label that specifies the meaning that Everything Has Taken As A Whole.

Although the question of whether you would like to live in such a world is an intelligible one, I think it misses the point entirely. For although some of us like to act as though some of the things we either love or hate have a meaning that is both unassailable and independent of us, this idea cannot withstand scrutiny. My meaning label world is meant to demonstrate this failure by bringing this notion front and center. Hopefully, the problem came screaming off the page as you read the previous paragraph in the form of the following question: But what if one of the labels I find does not reflect what this person, place, or thing means *to me*? Choosing what something means to us must certainly be at or very near the top of the list of things we are free to do if we are free in even the most minimal sense.

As it turns out, this fact is directly in line with the conclusion of my main project. Frankfurt (2004: 23) returns to this theme time and again in claims like, "It is by caring about things that we infuse the world with importance." This idea is very closely related to a central theme of existentialism. As Jean-

Paul Sartre (1969: 566) says, "life has no meaning *a priori*. . . . It is up to you to give it a meaning, and value is nothing but the meaning you choose." There are no exceptions to this rule for Sartre (and probably not for Frankfurt either), "not even a valid proof of the existence of God" (Sartre 2005: 214). This is because, as the meaning label case was meant to show, you would still have to decide what this meant, if anything, to you.

I could elaborate on these points further, but one of the most enjoyable—and refreshingly brief—works of philosophy I have encountered is on this very subject. When I realized that my project squared nicely with Thomas Nagel's (1987: 95-101) essay called *The Meaning of Life*, I knew I had to include it rather than attempt a feeble summary:

> Perhaps you have had the thought that nothing really matters, because in two hundred years we'll all be dead. This is a peculiar thought, because it's not clear why the fact that we'll be dead in two hundred years should imply that nothing we do now really matters.
>
> The idea seems to be that we are in some kind of rat race, struggling to achieve our goals and make something of our lives, but that this makes sense only if those achievements will be permanent. But they won't be. Even if you produce a great work of literature which continues to be read thousands of years from now, eventually the solar system will cool or the universe will wind down or collapse, and all trace of your efforts will vanish. In any case, we can't hope for even a fraction of this sort of immortality. If there's any point at all to what we do, we have to find it within our own lives.
>
> Why is there any difficulty in that? You can explain the point of most of the things you do. You work to earn money to support yourself and perhaps your family. You eat because you're hungry, sleep because you're tired, go for a walk or call up a friend because you feel like it, read the newspaper to find out what's going on in the world. If you didn't do any of those things you'd be miserable; so what's the big problem?

The problem is that although there are justifications and explanations for most of the things, big and small, that we do *within* life, none of these explanations explain the point of your life as a whole—the whole of which all these activities, successes and failures, strivings and disappointments are parts. If you think about the whole thing, there seems to be no point to it at all. Looking at it from the outside, it wouldn't matter if you had never existed. And after you have gone out of existence, it won't matter that you did exist.

Of course your existence matters to other people—your parents and others who care about you—but taken as a whole, their lives have no point either, so it ultimately doesn't matter that you matter to them. You matter to them and they matter to you, and that may give your life a feeling of significance, but you're just taking in each other's washing, so to speak. Given that any person exists, he has needs and concerns which make particular things and people within his life matter to him. But the *whole thing* doesn't matter.

But does it matter that it doesn't matter? "So what?" you might say. "It's enough that it matters whether I get to the station before my train leaves, or whether I've remembered to feed the cat. I don't need more than that to keep going." This is a perfectly good reply. But it only works if you really can avoid setting your sights higher, and asking what the point of the whole thing is. For once you do that, you open yourself to the possibility that your life is meaningless.

The thought that you'll be dead in two hundred years is just a way of seeing your life embedded in a larger context, so that the point of smaller things inside it seems not to be enough—seems to leave a larger question unanswered. But what if your life as a whole did have a point in relation to something larger? Would that mean that it wasn't meaningless after all?

There are various ways your life could have a larger meaning. You might be part of a political or social movement which changed the world for the better, to the benefit of future generations. Or you might just help provide a good life for your own children and their descendants. Or your life might be thought to have meaning in a religious context, so that your time on Earth was just a preparation for an eternity in direct contact with God.

About the types of meaning that depend on relations to other people, even people in the distant future, I've already indicated what the problem is. If one's life has a point as a part of something larger, it is still possible to ask about that larger thing, what is the point of *it*? Either there's an answer in terms of something still larger or there isn't. If there is, we simply repeat the question. If

there isn't, then our search for a point has come to an end with something which has no point. But if that pointlessness is acceptable for the larger thing of which our life is a part, why shouldn't it be acceptable already for our life taken as a whole? Why isn't it all right for your life to be pointless? And if it isn't acceptable there, why should it be acceptable when we get to the larger context? Why don't we have to go on to ask, "But what is the point of all *that*?" (human history, the succession of the generations, or whatever).

The appeal to a religious meaning to life is a bit different. If you believe that the meaning of your life comes from fulfilling the purpose of God, who loves you, and seeing Him in eternity, then it doesn't seem appropriate to ask, "And what is the point of *that*?" It's supposed to be something which is its own point, and can't have a purpose outside itself. But for this very reason it has its own problems.

The idea of God seems to be the idea of something that can explain everything else, without having to be explained itself. But it's very hard to understand how there could be such a thing. If we ask the question, "Why is the world like this?" and are offered a religious answer, how can we be prevented from asking again, "And why is *that* true?" What kind of answer would bring all of our "Why?" questions to a stop, once and for all? And if they can stop there, why couldn't they have stopped earlier?

The same problem seems to arise if God and His purposes are offered as the ultimate explanation of the value and meaning of our lives. The idea that our lives fulfil God's purpose is supposed to give them their point, in a way that doesn't require or admit of any further point. One isn't supposed to ask "What is the point of God?" any more than one is supposed to ask, "What is the explanation of God?"

But my problem here, as with the role of God as ultimate explanation, is that I'm not sure I understand the idea. Can there really be something which gives point to everything else by encompassing it, but which couldn't have, or need, any point itself? Something whose point can't be questioned from outside because there is no outside?

If God is supposed to give our lives a meaning that we can't understand, it's not much of a consolation. God as ultimate justification, like God as ultimate explanation, may be an incomprehensible answer to a question that we can't get rid of. On the other hand, maybe that's the whole point, and I am just failing to understand religious ideas. Perhaps the belief in God is the belief that the universe is intelligible, but not to us.

Leaving that issue aside, let me return to the smaller-scale dimensions of human life. Even if life as a whole is meaningless,

> perhaps that's nothing to worry about. Perhaps we can recognize it and just go on as before. The trick is to keep your eyes on what's in front of you, and allow justifications to come to an end inside your life, and inside the lives of others to whom you are connected. If you ever ask yourself the question, "But what's the point of being alive at all?"—leading the particular life of a student or bartender or whatever you happen to be—you'll answer "There's no point. It wouldn't matter if I didn't exist at all, or if I didn't care about anything. But I do. That's all there is to it."
>
> Some people find this attitude perfectly satisfying. Others find it depressing, though unavoidable. Part of the problem is that some of us have an incurable tendency to take ourselves seriously. We want to matter to ourselves "from the outside." If our lives as a whole seem pointless, then a part of us is dissatisfied—the part that is always looking over our shoulders at what we are doing. Many human efforts, particularly those in the service of serious ambitions rather than just comfort and survival, get some of their energy from a sense of importance—a sense that what you are doing is not just important to you, but important in some larger sense: important, period. If we have to give this up, it may threaten to take the wind out of our sails. If life is not real, life is not earnest, and the grave is its goal, perhaps it's ridiculous to take ourselves so seriously. On the other hand, if we can't help taking ourselves so seriously, perhaps we just have to put up with being ridiculous. Life may be not only meaningless but absurd.

Count me as one who finds this answer perfectly satisfying. After all, if there is no other plausible[1] answer, then what is the point of fretting about it?

Leaving that to one side, I have a final point to make about Nagel's piece as it relates to my project. In the search for meaning within your own life, the external search should be paired with a matching internal search. The internal search is necessary because, given our particular predilections and proclivities, there will be some things that we are *able* to care about and other things that

---

[1] I do not think this word is strong enough to describe the situation. As Nagel says, even though some may find it depressing, they still find it unavoidable. So it may not be that other options are *implausible*, as much as it is that they may be difficult or impossible to describe coherently.

we are incapable of caring about (Frankfurt 1999: 178-79).[2] It is also worth noting that the facts concerning what we are able to care about may not line up with our preferred conception of ourselves. And there is more likely to be a discrepancy here when it comes to the people we care about. Our preferred conception of them may have no overlap at all with what they are in fact capable of caring about. The most benign manifestation of forgetting this fact is an argument among adults. The most tragic is the parents who attempt to force the things they care about on their children.

Of course, the scarring from this practice can range from mildly amusing (as portrayed nicely in the film *The Breakfast Club*) all the way down to utterly horrific (as with the "honor" killings of homosexual or otherwise sexually active children). So parents, before indiscriminately inflicting your values on your children, remember that you are potentially costing the people you care about a chance of leading a life filled with things *they* can and do care about. Imagine a world where well-meaning parents ensured the existence of Aristotle the carpenter, Mozart the doctor, and Van Gogh the lawyer. Or just totally ignore this point. You know what is best for your kid, right? And if you mess it up, you can always take comfort in the fact that the sun will eventually swallow the earth—erasing all evidence of you and your mistake—before the universe either goes cold or collapses.

Have a nice day. ☺

---

[2] Rosati (1995: 300 n.10) also discusses the claim that "it is a necessary condition on something being good for a person that she be capable of caring about it."

BIBLIOGRAPHY

Adams, Robert, *Finite and Infinite Goods* (New York: Oxford University Press, 1999).

Alston, W. P., "Pleasure," in Paul Edwards (ed.) *The Encyclopedia of Philosophy* (New York: Macmillan, 1967): 341-347.

Aristotle, *Nicomachean Ethics.*

Aristotle, *Metaphysics.*

Arneson, R.J., "Human Flourishing Versus Desire Satisfaction," *Social Philosophy and Policy* 16 (1999): 113-142.

Boonin, David, *The Problem of Punishment* (Cambridge, UK: Cambridge University Press, 2008).

Bradley, Ben, "A Paradox for Some Theories of Welfare," *Philosophical Studies* 133 (2007): 45-53.

Brandt, Richard, "Rationality, Egoism, and Morality," *The Journal of Philosophy* 69 (1972): 681-698.

Brandt, Richard, *A Theory of the Good and the Right* (Oxford: Oxford University Press, 1979).

Brandt, Richard, "Two Concepts of Utility," in Miller and Williams (eds.) *The Limits of Utilitarianism* (Minneapolis, MN: University of Minnesota Press, 1982): 169-185.

Broad, C. D., *Five Types of Ethical Theory* (New York: Harcourt, Brace and Co., 1930).

Camus, Albert, "The Myth of Sisyphus," in Oaklander (ed.) *Existentialist Philosophy: An Introduction* (Upper Saddle River, NJ: Prentice-Hall, Inc., 1996): 357-369.

Carson, Thomas L., *Value and the Good Life* (Notre Dame, IN: University of Notre Dame Press, 2000).

Chisholm, Roderick, "Freedom of Action," in K. Lehrer (ed.) *Freedom and Determinism* (New York: Random House, 1966): 11-44.

Darwall, Stephen, *Impartial Reason* (Ithaca, NY: Cornell University Press, 1983).

Darwall, Stephen, *Welfare and Rational Care* (Princeton, NJ: Princeton University Press, 2002).

Davidson, Donald, *Essays on Actions and Events* (Oxford: Oxford University Press, 2001).

Davis, Wayne A., "The Two Senses of Desire," *Philosophical Studies* 45 (1984): 181-195.

Descartes, *Meditations on First Philosophy* (Cambridge, UK: Cambridge University Press, 1996).

Dostoevsky, Fyodor, *Notes from Underground* (New York: W. W. Norton & Company, 1989).

Fehige, Christoph and Ulla Wessels (eds.), *Preferences* (Berlin and New York: Walter de Gruyter, 1998).

Feinberg, Joel, "Psychological Egoism," in S. Cahn and P. Markie (eds.) *Ethics: History, Theory, and Contemporary Issues* (New York: Oxford University Press, 2006): 527-534.

Feldman, Fred, "Two Questions about Pleasure," in F. Feldman (ed.) *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy* (Cambridge, UK: Cambridge University Press, 1997): 82-105.

Feldman, Fred, "The Good Life: A Defense of Attitudinal Hedonism," *Philosophy and Phenomenological Research* 65 (2002): 604-628.

Feldman, Fred, *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism* (New York: Oxford University Press, 2004).

Frankena, William, *Ethics* (Englewood Cliffs, NJ: Prentice-Hall, Inc., 1973).

Frankfurt, Harry G., *The Importance of What We Care About* (Cambridge, UK: Cambridge University Press, 1988).

Frankfurt, Harry G., *Necessity, Volition, and Love* (Cambridge, UK: Cambridge University Press, 1999).

Frankfurt, Harry G., *The Reasons of Love* (Princeton, NJ: Princeton University Press, 2004).

Frankfurt, Harry G., *Taking Ourselves Seriously and Getting It Right* (Stanford, CA: Stanford University Press, 2006).

Fuchs, Alan E., "Posthumous Satisfactions and the Concept of Individual Welfare," in J. Heil (ed.) *Rationality, Morality, and Self-Interest: Essays Honoring Mark Carl Overvold* (Lanham, MD: Rowman & Littlefield, 1993): 215-220.

Gibbard, Allan, "Ordinal Utilitarianism," in G.R. Feiwel (ed.) *Arrow and the Foundations of the Theory of Economic Policy* (New York: New York University Press, 1987): 135–153.

Gibbard, Allan, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, MA: Harvard University Press, 1990).

Griffin, James, *Well-Being* (Oxford: Oxford University Press, 1986).

Haji, Ishtiyaque, *Freedom and Value* (Springer, 2009).

Hare, R. M., *Moral Thinking: Its Levels, Method, and Point* (Oxford: Oxford University Press, 1981).

Haslett, D. W., "What Is Utility?", *Economics and Philosophy* 6 (1990): 65-94.

Haybron, Daniel M., *The Pursuit of Unhappiness: The Elusive Psychology of Well-Being* (Oxford: Oxford University Press, 2008).

Heathwood, Chris, "The Problem of Defective Desires," *Australasian Journal of Philosophy* 83 (2005): 487-504.

Heathwood, Chris, "Desire Satisfactionism and Hedonism," *Philosophical Studies* 128 (2006): 539-563.

Heathwood, Chris, "The Reduction of Sensory Desire to Pleasure," *Philosophical Studies* 133 (2007): 23-44.

Heathwood, Chris, "Preferentism and Self-Sacrifice," *Pacific Philosophical Quarterly* 92 (2011a): 18–38.

Heathwood, Chris, "Desire-Based Theories of Reasons, Pleasure, and Welfare," in Russ Shafer-Landau (ed.) *Oxford Studies in Metaethics*, Vol. 6 (Oxford: Oxford University Press, 2011b): 79-106.

Heathwood, Chris, *Subjective Desire Satisfactionism*, Unpublished Manuscript.

Hobbes, Thomas, *Leviathan* (Broadview Press, 2002).

Huemer, Michael, *Ethical Intuitionism* (New York: Palgrave Macmillan, 2005).

Hume, David, *An Enquiry Concerning Human Understanding* (Indianapolis, IN: Hackett Publishing Co., 1993).

Hume, David, *A Treatise of Human Nature* (Mineola, NY: Dover Publications, Inc., 2003).

Hume, David, *Dialogues Concerning Natural Religion* (Amherst, NY: Prometheus Books, 1989).

James, William, "The Moral Philosopher and the Moral Life," in Alburey Castell (ed.) *Essays in Pragmatism by William James* (New York: Hafner Publishing Co., 1948): 65-87.

Kagan, Shelly, "The Limits of Well-Being," *Social Philosophy and Policy* 9 (1992): 169-189.

Kagan, Shelly, *Normative Ethics* (Boulder, CO: Westview Press, 1998).

Keller, Simon, "Welfare and the Achievement of Goals," *Philosophical Studies* 121 (2004): 27-41.

Kierkegaard, Søren, *The Essential Kierkegaard* (Princeton, NJ: Princeton University Press, 2000).

Kim, Jaegwon, *Philosophy of Mind* (Westview Press, 2006).

Kraut, Richard, "Desire and the Human Good," *Proceedings and Addresses of the American Philosophical Association* 68 (1994): 39-54.

Lewis, David, "Desire as Belief," *Mind* 97 (1988): 323-332.

Lewis, David, "Dispositional Theories of Value," *Proceedings of the Aristotelian Society* Supp. Vol. 63 (1989): 113-137.

Loeb, Don, "Full-Information Theories of Individual Good," *Social Theory and Practice* 21 (1995): 1-30.

Lonsky, Loren E., "Person, Concept Of," in Becker and Becker (eds.) *Encyclopedia of Ethics*, Vol. III (New York, NY: Routledge, 2001): 1293.

Mill, John Stuart, "Utilitarianism," in S. Cahn and P. Markie (eds.) *Ethics: History, Theory, and Contemporary Issues* (New York: Oxford University Press, 2006): 317-350.

Moore, G. E., *Principia Ethica* (Cambridge, UK: Cambridge University Press, 1903).

Nagel, Thomas, "Death," *Noûs* 4 (1970): 73-80.

Nagel, Thomas, "What Is It Like to Be a Bat?", *The Philosophical Review* 83 (1974): 435-450.

Nagel, Thomas, *What Does It All Mean?: A Very Short Introduction to Philosophy* (New York: Oxford University Press, 1987).

Nietzsche, Friedrich, *Beyond Good and Evil* (New York: Vintage Books, 1966).

Nietzsche, Friedrich, *Human, All Too Human: A Book for Free Spirits* (Cambridge, UK: Cambridge University Press, 1996).

Nietzsche, Friedrich, *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings* (Cambridge, UK: Cambridge University Press, 2005).

Noggle, Robert, "Integrity, the Self, and Desire-Based Accounts of the Good," *Philosophical Studies* 96 (1999): 303-331.

Nozick, Robert, *Anarchy, State, & Utopia* (New York: Basic Books, Inc., 1974).

Nussbaum, Martha, *Sex & Social Justice* (New York: Oxford University Press, 1999).

Overvold, Mark Carl, "Self-Interest and the Concept of Self-Sacrifice," *Canadian Journal of Philosophy* 10 (1980): 105-118.

Overvold, Mark Carl, "Self-Interest and Getting What You Want," in Miller and Williams (eds.) *The Limits of Utilitarianism* (Minneapolis, MN: University of Minnesota Press, 1982): 186-194.

Parfit, Derek, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

Parfit, Derek, "Overpopulation and the Quality of Life" in Peter Singer (ed.) *Applied Ethics* (Oxford: Oxford University Press, 1986): 145-164.

Penelhum, Terence, "The Importance of Self-Identity," *Journal of Philosophy* 68 (1971): 667-678.

Pereboom, Derk, "Free Will," in Becker and Becker (eds.) *Encyclopedia of Ethics*, Vol. I (New York: Routledge 2001): 571-574.

Plato, *Apology*.

Plato, *Philebus*.

Portmore, Douglas, "Desire Fulfillment and Posthumous Harm," *American Philosophical Quarterly* 44 (2007): 27-38.

Railton, Peter, "Facts and Values," *Philosophical Topics* 14 (1986): 5-31.

Railton, Peter, *Facts, Values, and Norms: Essays Toward a Morality of Consequence* (Cambridge, UK: Cambridge University Press, 2003).

Rawls, John, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1999).

Regan, Tom, "Animals, Treatment Of," in Becker and Becker (eds.) *Encyclopedia of Ethics*, Vol. III (New York: Routledge, 2001): 71.

Rosati, Connie, "Persons, Perspectives, and Full Information Accounts of the Good," *Ethics* 105 (1995): 296-325.

Ross, W. D., *The Right and the Good* (Hackett Publishing Co., 1930).

Rundle, Bede, *Why There Is Something Rather Than Nothing* (Oxford: Oxford University Press, 2004).

Russell, Bertrand, *The Problems of Philosophy* (Amherst, NY: Prometheus Books, 1988).

Saint Augustine, *Confessions*.

Sartre, Jean-Paul, *Being and Nothingness: An Essay on Phenomenological Ontology* (New York: Routledge, 1969).

Sartre, Jean-Paul, "Existentialism Is a Humanism," in Robert Solomon (ed.) *Existentialism* (New York: Oxford University Press, 2005): 206-214.

Scanlon, T. M., *What We Owe to Each Other* (Cambridge, MA: Harvard University Press, 1998).

Schueler, G. F., *Desire: Its Role in Practical Reason and the Explanation of Action* (Cambridge, MA: The MIT Press, 1995).

Schwartz, Thomas, "Human Welfare: What It Is Not," in Miller and Williams (eds.) *The Limits of Utilitarianism* (Minneapolis, MN: University of Minnesota Press, 1982): 195-206.

Sen, Amartya, "Rational Fools: A Critique of the Behavioral Foundations of Economic Theory," *Philosophy and Public Affairs* 6 (1977): 317-344.

Sen, Amartya, "Well-Being, Agency and Freedom: The Dewey Lectures 1984," *The Journal of Philosophy* 82 (1985): 169-221.

Sidgwick, Henry, *The Methods of Ethics* (Macmillan, 1907).

Singer, Peter, "All Animals Are Equal," in Regan and Singer (eds.), *Animal Rights and Human Obligations* (New Jersey: Prentice-Hall, 1989): 148-162.

Sobel, David, "Full Information Accounts of Well-Being," *Ethics* 104 (1994): 784-810.

Sobel, David, "Varieties of Hedonism," *Journal of Social Philosophy* 33 (2002): 240-256.

Spinoza, Baruch, *The Ethics* (Joseph Simon, 1981).

Sumner, L. W., *Welfare, Happiness, & Ethics* (Oxford: Oxford University Press, 1996).

Thomson, Judith Jarvis, "A Defense of Abortion," *Philosophy and Public Affairs* 1 (1971): 47-66.

Tooley, Michael, "Abortion and Infanticide," *Philosophy and Public Affairs* 2 (1972): 37-65.

Tooley, Michael, "The Problem of Evil," in Zalta (ed.) *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), URL = <http://plato.stanford.edu/archives/spr2010/entries/evil/>.

Velleman, J. David, "Brandt's Definition of 'Good,'" *The Philosophical Review* 97 (1988): 353-371.

Velleman, J. David, "Well-Being and Time," *Pacific Philosophical Quarterly* 72 (1991): 48-77.

Wertheimer, Roger, "Understanding the Abortion Argument," *Philosophy & Public Affairs* 1 (1971): 67-95.